

INVARIANCE PRINCIPLES FOR TREE-VALUED CANNINGS CHAINS

STEPHAN GUFLER

ABSTRACT. We consider sequences of tree-valued Markov chains that describe evolving genealogies in Cannings models, and we show their convergence in distribution to tree-valued Fleming-Viot processes. Under the conditions of Möhle and Sagitov, this convergence holds for all tree-valued Fleming-Viot processes under consideration in the dust-free case, and for the Fleming-Viot processes with values in the space of distance matrix distributions in the case with dust. Convergence to Fleming-Viot processes with values in the space of marked metric measure spaces in the case with dust is ensured by an additional assumption on the probability that a randomly sampled individual belongs to a non-singleton family.

CONTENTS

1. Introduction	1
2. Preliminaries	3
2.1. Distance matrices	3
2.2. Metric measure spaces and marked metric measure spaces	4
3. Invariance principles	6
3.1. The Cannings model and the process of the genealogical distances	6
3.2. Processes with values in the space of metric measure spaces	7
3.3. Processes of distance matrix distributions	8
3.4. Processes with values in the space of marked metric measure spaces	8
3.5. Another decomposition of the evolving genealogical trees	9
4. Tree-valued Fleming-Viot processes	11
4.1. The \mathbb{U} -valued Ξ -Fleming-Viot process	11
4.2. The $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process	12
4.3. The \mathcal{U}^{erg} -valued Ξ -Fleming-Viot process	13
5. An example	13
6. Convergence of the transition kernels	15
6.1. Proofs of invariance principles	15
6.2. Proofs of Lemmas 6.1 and 6.2	18
7. Convergence of marked metric measure spaces in the dust-free case	21
8. Convergence of marked metric measure spaces in the case with dust	24
List of notation	29
References	30

1. INTRODUCTION

In population genetics, Cannings models [5, 6] are classical evolutionary models with constant population size N and non-overlapping generations. Reproduction events occur independently between the generations such that the individuals in generation $k+1$ are subdivided into families according to an exchangeable random partition and each family draws its ancestor in generation k independently without replacement. The Wright-Fisher model is a classical example for a Cannings model. In the Wright-Fisher model, each individual in generation $k+1$ draws its ancestor in generation k independently with replacement. We refer the reader also to e. g. [10] for these models.

2010 *Mathematics Subject Classification.* Primary 60F17, Secondary 60J25, 60K35, 92D10.

Key words and phrases. Invariance principle, Cannings model, tree-valued Fleming-Viot process, (marked) metric measure space, (marked) Gromov-Prohorov metric, external branches.

Coalescents with simultaneous multiple mergers are robust infinite population size limits of partition-valued processes that describe the genealogies in Cannings models at fixed times. Möhle and Sagitov [22] give a criterion for this convergence. Sagitov [24] gives an equivalent criterion in terms of the measure Ξ of Schweinsberg [25]. A first robustness result for the Kingman coalescent is shown in Kingman [17], see also e.g. [2, Theorem 2].

In the present article, we consider evolving genealogies, and we describe the genealogical tree of the individuals in Cannings models at each time in various ways. First, we consider the metric measure space that consists of the set of individuals at that time, their mutual genealogical distances, and the uniform probability measure. Second, we consider the distance matrix distribution of the aforementioned metric measure space, i.e. the distribution of the infinite matrix of the genealogical distances between iid samples. Third, we decompose the genealogical tree into the external branches and the remaining subtree. We then consider the semi-metric space that consists of the starting vertices of the external branches and their mutual distances. On the product space of this semi-metric space with \mathbb{R}_+ , we define a probability measure such that for each external branch, mass $1/N$ is added to the pair that consists of its starting vertex and its length. We then obtain a marked metric measure space. Fourth, we also decompose the genealogical tree and describe it by a marked metric measure space, pruning not the whole external branch, but only the part from each leaf to the most recent reproducing individual on the ancestral lineage.

We endow the space of distance matrix distributions with the Prohorov metric, and the space of (marked) metric measure spaces with the (marked) Gromov-Prohorov metric [8, 12]. We consider Markov chains whose states describe the evolving genealogy in Cannings models in one of the four aforementioned ways. Our invariance principles show that sequences of such Markov chains converge in distribution, under a rescaling of time and genealogical distances, in the space of càdlàg paths in the respective state space, endowed with the Skorohod topology. The limit processes are the Fleming-Viot processes from [16, Section 8]. Under the condition of Sagitov [24] and the assumption that the initial states converge, we show the convergence of the prelimiting chains with values in the space of distance matrices. For the convergence of the prelimiting chains with values in the space of metric measure spaces, we have to assume in addition that the limiting genealogy is dust-free as the tree-valued Fleming-Viot process with values in the space of metric measure spaces exists only in the dust-free case. Dust-freeness can be characterized by the property that a randomly drawn external branch has a.s. length zero (cf. Propositions 6.6 and 7.4 in [16]).

In the dust-free case, the sequences of prelimiting chains for the third and the fourth description of the genealogy converge under the condition of Sagitov [24] and an appropriate condition on the initial state. We show their convergence in the case with dust under the additional assumption (3.8) on the probability that a randomly sampled individual from a fixed generation belongs to a non-singleton family. An additional assumption is needed here as the convergence in the marked Gromov-Prohorov metric (other than the weak convergence of the distance matrix distributions) implies weak convergence of the empirical distribution of the external branch lengths or the distances to the most recent reproducing individual, respectively.

In Section 2.1, we recall the decomposition of the genealogical trees at the external branches. In Section 2.2, we recall some notions on metric measure spaces and marked metric measure spaces. We state our convergence results in Section 3. We recall tree-valued Ξ -Fleming-Viot processes in Section 4. In Section 5, we give an example in which assumption (3.8) is not satisfied and the chains with values in the space of marked metric measure spaces do not converge. The proofs of the invariance principles are given in the further sections.

The tree-valued Fleming-Viot process with binary reproduction events is introduced in Greven, Pfaffelhuber, and Winter [13] as the solution of a well-posed martingale problem that is the limit in distribution of tree-valued processes read off from Moran models. In [13, Remark 2.21], it is conjectured that a tree-valued Fleming-Viot process is the robust limit of tree-valued processes read off from Cannings models. In [16], tree-valued Fleming-Viot processes are studied in the setting with simultaneous multiple reproduction events, the case with dust is included by the

decomposition of the genealogical trees into the external branches and the remaining subtree. These decomposed genealogical trees are described by marked metric measure spaces and their distance matrix distributions. Path regularity of tree-valued Fleming-Viot processes follows from the pathwise construction in [15].

In Section 6, we prove the invariance principles for the Markov chains associated with the first, second, and in the case with dust also for the fourth of the above descriptions of the genealogy. Here we can apply a general convergence result from Ethier and Kurtz [11, Chapter 4.8] as the transition kernels of the prelimiting chains converge on a core (in the sense of [11, Chapter 1.3]) to the generators of the tree-valued Fleming-Viot processes. In [16], it is shown that the domains of the martingale problems for the tree-valued Fleming-Viot processes are cores, and that the semigroups on these cores are strongly continuous. We use existence and path regularity of the limit processes, and we do not need to show relative compactness of the prelimiting processes here.

We prove the invariance principles for the processes with values in the space of marked metric measure spaces in Sections 7 and 8 by comparison with processes whose convergence is proved in Section 6. In the case without dust, we can compare these process in the supremum metric. In the case with dust, we compare only the finite dimensional distributions whence we also have to check then the relative compactness of the prelimiting processes.

We use also exchangeable random partitions, Kingman's correspondence and its continuity properties, for which we refer to Pitman [23, Chapter 2] and Bertoin [1, Chapter 2]. We mention that Stournaras [26] uses [11, Corollary 4.8.17] to show, by verifying the compact containment condition via [13, Proposition 2.22] with some effort, convergence of tree-valued Wright-Fisher models to the tree-valued Fleming-Viot process from [13].

2. PRELIMINARIES

It is well-known that ultrametric spaces can be viewed as leaf-labeled real trees (cf. e. g. [16, Remark 1.1]). In Subsection 2.1, we recall a decomposition of semi-ultrametrics that corresponds to the decomposition of the associated trees at the external branches. In Subsection 2.2, we recall isomorphism classes of metric measure spaces and marked metric measure spaces which we can be interpreted as unlabeled genealogical trees. In the finite case, the (marked) metric measure space associated with a (decomposed) ultrametric can be viewed as the equivalence class under permutations of the labels of the leaves.

2.1. Distance matrices. We denote the set of the positive integers by \mathbb{N} , the set of the non-negative integers by \mathbb{N}_0 , and for $N \in \mathbb{N}$, we write $[N] = \{1, \dots, N\}$. Let \mathfrak{D} denote the space of semimetrics on \mathbb{N} and $\mathfrak{U} \subset \mathfrak{D}$ the set of semi-ultrametrics on \mathbb{N} . We do not distinguish between a semi-metric $\rho \in \mathfrak{D}$ and the distance matrix $(\rho(i, j))_{i, j \in \mathbb{N}}$, and we view \mathfrak{U} and \mathfrak{D} as subspaces of the space $\mathbb{R}^{\mathbb{N}^2}$ which we endow with a complete and separable metric that induces the product topology, where \mathbb{R} is endowed with the Euclidean topology. Analogously, for $N \in \mathbb{N}$, we denote by \mathfrak{D}_N the space of semimetrics on $[N]$ and by $\mathfrak{U}_N \subset \mathfrak{D}_N$ the space of semi-ultrametrics on $[N]$. Again we do not distinguish between semi-metrics and distance matrices and we view \mathfrak{D}_N as a subspace of the finite-dimensional space \mathbb{R}^{N^2} which we endow with a norm and the (induced) Euclidean topology.

We now decompose semi-ultrametrics in \mathfrak{U}_N as in [16, Section 2]. The continuous map

$$\alpha : \mathbb{R}_+^{N^2} \times \mathbb{R}_+^N \rightarrow \mathbb{R}_+^{N^2}, \quad (r, v) \mapsto ((v(i) + r(i, j) + v(j)) \mathbf{1}\{i \neq j\})_{i, j \in [N]},$$

is used to retrieve ultrametric distance matrices from the elements of the space

$$\hat{\mathfrak{U}}_N = \{(r, v) \in \mathfrak{D}_N \times \mathbb{R}_+^N : \alpha(r, v) \in \mathfrak{U}_N\}$$

of decomposed semi-ultrametrics which we also call marked distance matrices. Conversely, for $N \geq 2$, we decompose ultrametric distance matrices using the maps

$$\Upsilon : \mathfrak{U}_N \rightarrow \mathbb{R}_+^N, \quad \rho \mapsto \left(\frac{1}{2} \min_{j \in [N] \setminus \{i\}} \rho(i, j)\right)_{i \in [N]}$$

and $\beta : \mathfrak{U}_N \rightarrow \hat{\mathfrak{U}}_N$, $\rho \mapsto (r, v)$, where $v = \Upsilon(\rho)$ and $r(i, j) = (\rho(i, j) - v(i) - v(j)) \mathbf{1}\{i \neq j\}$ for $i, j \in [N]$. For $\rho \in \mathfrak{U}_1$, we set $\Upsilon(\rho) = 0$ and $\beta(\rho) = (\rho, 0)$.

As in [16, Remark 2.2], the quantity $v(i)$ is the length of the external branch that ends in leaf i of the coalescent tree associated with ρ , and $r(i, j)$ is the distance between the starting vertices of the external branches that end in leaves i and j , respectively. Here an external branch is defined to consist only of the leaf i if there exists $j \in [N] \setminus \{i\}$ with $\rho(i, j) = 0$. (In fact, the finite setting discussed in this subsection can be seen as a special case of Section 2 in [16] as any semi-ultrametric ρ on $[N]$ can be extended to \mathbb{N} by setting e.g. $\rho(1, k) = 0$ for $k > N$.)

We also use the decomposition of semi-ultrametrics in \mathfrak{U} from [16, Section 2]. Here we have the continuous map

$$\alpha : \mathbb{R}_+^{\mathbb{N}^2} \times \mathbb{R}_+^{\mathbb{N}} \rightarrow \mathbb{R}_+^{\mathbb{N}^2}, \quad (r, v) \mapsto ((v(i) + r(i, j) + v(j)) \mathbf{1}\{i \neq j\})_{i, j \in \mathbb{N}},$$

the space $\hat{\mathfrak{U}} = \{(r, v) \in \mathfrak{D} \times \mathbb{R}_+^{\mathbb{N}} : \alpha(r, v) \in \mathfrak{U}\} \subset \mathbb{R}_+^{\mathbb{N}^2} \times \mathbb{R}_+^{\mathbb{N}}$ of marked distance matrices (or decomposed semi-ultrametrics), and the map

$$\Upsilon : \mathfrak{U} \rightarrow \mathbb{R}_+^{\mathbb{N}}, \quad \rho \mapsto \left(\frac{1}{2} \inf_{j \in \mathbb{N} \setminus \{i\}} \rho(i, j)\right)_{i \in \mathbb{N}}$$

which maps a semi-ultrametric ρ to the sequence of the external branch lengths of the associated tree.

2.2. Metric measure spaces and marked metric measure spaces. For the theory of metric measure spaces and marked metric measure spaces, we refer to [8, 12, 14, 19, 27].

A metric measure space is a triple (X, ρ, μ) that consists of a complete and separable metric space (X, ρ) and a probability measure μ on the Borel sigma algebra on X . Two metric measure spaces (X, ρ, μ) and (X', ρ', μ') are defined to be isomorphic if there exists a measure-preserving isometry between the supports $\text{supp } \mu$ and $\text{supp } \mu'$. The distance matrix distribution $\nu^{(X, \rho, \mu)}$ of a metric measure space (X, ρ, μ) is defined as the distribution of the random matrix $(\rho(x(i), x(j)))_{i, j \in \mathbb{N}}$, where $(x(i), i \in \mathbb{N})$ is a μ -iid sequence in X . The Gromov reconstruction theorem ([27, Theorem 4] and [14, Section 3 $\frac{1}{2}$]) states that metric measure spaces are isomorphic if and only if they have the same distance matrix distribution. Hence, we can define the isomorphism class $\llbracket X, \rho, \mu \rrbracket$ of a metric measure space (X, ρ, μ) such that

$$\mathbb{U} = \{\llbracket X', \rho', \mu' \rrbracket : (X', \rho', \mu') \text{ ultrametric measure space}\}$$

is a set. For $\chi \in \mathbb{U}$, we denote the associated distance matrix distribution by ν^χ . A sequence $(\chi_k, k \in \mathbb{N})$ in \mathbb{U} converges to $\chi \in \mathbb{U}$ in the Gromov-weak topology if and only if the distance matrix distributions ν^{χ_k} converge weakly to ν^χ . The Gromov-Prohorov metric d_{GP} induces the Gromov-weak topology and is complete and separable, as shown in [12]. As in [13], the elements of \mathbb{U} can be considered as trees.

For $N \in \mathbb{N}$, we also work with the closed subspace

$$\mathbb{U}_N = \{\llbracket X, \rho, \mu \rrbracket : (X, \rho, \mu) \text{ ultrametric measure space such that } N\mu \text{ is integer-valued}\}$$

of \mathbb{U} which can be interpreted as the space of semi-ultrametric spaces that contain N elements and that are endowed with the uniform probability measure. When we identify points $x, y \in X$ with $\rho(x, y) = 0$ in a semi-metric space (X, ρ) to obtain a metric space, we refer by x, y also to the corresponding element of the metric space, in slight abuse of notation. We define the isomorphism class of a semi-metric measure space as the isomorphism class of the metric measure space obtained by identifying the elements with distance zero. We define the function

$$\psi_N : \mathfrak{U}_N \rightarrow \mathbb{U}_N, \quad \rho \mapsto \llbracket [N], \rho, N^{-1} \sum_{i=1}^N \delta_i \rrbracket$$

which maps a semi-ultrametric to the isomorphism class of the associated semi-metric measure space with the uniform measure. It is clear that the map ψ_N is continuous, for formal proofs, cf. [16, Remark 11.1] or [15, Lemma 4.5].

For each $\chi \in \mathbb{U}_N$, there exists $\rho \in \mathfrak{U}_N$ with $\psi_N(\rho) = \chi$, and the N -distance matrix distribution $\nu^{N,\chi}$ is defined as the distribution of the random matrix $(\rho(x(i), x(j)))_{i,j \in [N]}$, where $x(1), \dots, x(N)$ are sampled from $[N]$ according to the uniform measure without replacement. (That is, $(x(i), i \in [N])$ is a uniform permutation of $[N]$.) For every \mathfrak{U}_N -valued random variable ρ' with distribution $\nu^{N,\chi}$, it holds $\psi_N(\rho') = \chi$ a.s. Hence, χ is uniquely determined by $\nu^{N,\chi}$, as in [13].

A (\mathbb{R}_+) -marked metric measure space is a triple (X, r, m) that consists of a complete and separable metric space (X, r) and a probability measure m on the Borel sigma algebra on the space $X \times \mathbb{R}_+$ which is endowed with the product metric $d((x, v), (x', v')) = r(x, x') \vee |v - v'|$. Two marked metric measure spaces (X, r, m) and (X', r', m') are defined to be isomorphic if there exists an isometry φ between the supports $\text{supp } m(\cdot \times \mathbb{R}_+)$ and $\text{supp } m'(\cdot \times \mathbb{R}_+)$ such that the isometry

$$\hat{\varphi} : \text{supp } m \rightarrow \text{supp } m', \quad (x, v) \mapsto (\varphi(x), v)$$

satisfies $\hat{\varphi}(m) = m'$. The marked distance matrix distribution $\nu^{(X, r, m)}$ of a marked metric measure space (X, r, m) is defined as the distribution of $((r(x(i), x(j))))_{i, j \in \mathbb{N}}, (v(i))_{i \in \mathbb{N}}$, where $((x(i), v(i)), i \in \mathbb{N})$ is an m -iid sequence in $X \times \mathbb{R}_+$. The Gromov reconstruction theorem for marked metric measure spaces (see [16, Proposition 3.12], [8, Theorem 1]) states that marked metric measure spaces are isomorphic if and only if they have the same marked distance matrix distribution. We can now define the isomorphism class $\llbracket X, r, m \rrbracket$ of a marked metric measure space (X, r, m) such that

$$\hat{\mathcal{U}} = \{\llbracket X', r', m' \rrbracket : (X', r', m') \text{ marked metric measure space with } \nu^{(X', r', m')}(\hat{\mathcal{U}}) = 1\}$$

is a set. As in [16], $\hat{\mathcal{U}}$ is the set of isomorphism classes of marked metric measure spaces that yield ultrametric spaces when marks in the support of the measure are added to the distances of the metric space, and the elements of $\hat{\mathcal{U}}$ can be viewed as trees.

We denote the marked distance matrix distribution associated with any $\chi \in \hat{\mathcal{U}}$ by ν^χ . Using the continuous map α from Subsection 2.1, we associate with χ the probability distribution $\alpha(\nu^\chi)$ on \mathfrak{U} which is called in [16] the distance matrix distribution of χ . (We denote by $f(\mu)$ the image measure of a measure μ under a map f). A sequence $(\chi_k, k \in \mathbb{N})$ converges to χ in $\hat{\mathcal{U}}$ in the marked Gromov-weak topology if and only if the marked distance matrix distributions ν^{χ_k} converge weakly to ν^χ . The marked Gromov-weak topology is metrized by the marked Gromov-Prohorov metric d_{mGP} which is complete and separable, see [8].

By [16, Proposition 3.3] each element of $\hat{\mathcal{U}}$ is uniquely characterized in $\hat{\mathcal{U}}$ by its distance matrix distribution. Hence, the set of distance matrix distributions of marked metric measure spaces, denoted (as in [16]) by

$$\mathcal{U}^{\text{erg}} := \{\alpha(\nu^\chi) : \chi \in \hat{\mathcal{U}}\},$$

is in one-to-one correspondence with $\hat{\mathcal{U}}$, and its elements can likewise be viewed as trees. We endow \mathcal{U}^{erg} with the Prohorov metric d_P . Then \mathcal{U}^{erg} is separable and by [16, Corollary 3.25] complete.

We say a marked metric measure (X, r, m) space supports only the zero mark if $m = \mu \otimes \delta_0$ for some probability measure μ on X . Clearly, this property depends only on the isomorphism class of the marked metric measure space. Also note that the distance matrix distribution of a marked metric measure space $(X, r, \mu \otimes \delta_0)$ that supports only the zero mark equals the distance matrix distribution of the metric measure space (X, r, μ) .

For $N \in \mathbb{N}$, we also define the closed subspace

$$\hat{\mathcal{U}}_N = \{\llbracket X, r, m \rrbracket \in \hat{\mathcal{U}} : (X, r, m) \text{ marked metric measure space, } Nm \text{ integer-valued}\}$$

of $\hat{\mathcal{U}}$ that stands for finite marked metric measure spaces. To obtain marked metric measure spaces from marked distance matrices, we use the map

$$\hat{\psi}_N : \hat{\mathfrak{U}}_N \rightarrow \hat{\mathcal{U}}_N, (r, v) \mapsto \llbracket [N], r, N^{-1} \sum_{i=1}^N \delta_{(i, v(i))} \rrbracket,$$

where we understand the isomorphy class of a marked semi-metric measure space as the isomorphy class of the marked metric measure space obtained by identifying elements of the metric space with distance zero. Clearly, the map $\hat{\psi}_N$ is continuous, a formal proof is given in [16, Remark 11.1].

As in [9], the N -marked distance matrix distribution of $\chi \in \hat{\mathcal{U}}_N$ is defined as the distribution of $((r(x(i), x(j)))_{i,j \in [N]}, (v(x(i)))_{i \in [N]})$, where (r, v) is any element of $\hat{\mathcal{U}}_N$ with $\hat{\psi}_N(r, v) = \chi$, and $x(1), \dots, x(N)$ is sampled uniformly from $[N]$ without replacement. Clearly, $\hat{\psi}_N(r', v') = \chi$ a.s. for any random variable (r', v') that has the marked distance matrix distribution of χ .

3. INVARIANCE PRINCIPLES

In Subsection 3.1, we define for each $N \in \mathbb{N}$ the Cannings population model of population size N . From this construction, we read off tree-valued processes in Subsections 3.2 – 3.5.

3.1. The Cannings model and the process of the genealogical distances. The population model has discrete generations, enumerated by \mathbb{N}_0 , and N individuals in each generation, labeled by $1, \dots, N$. The dynamics is characterized by a probability measure Ξ^N on the subspace

$$\Delta^N = \{x \in \Delta : |x|_1 = 1, Nx(i) \in \mathbb{N}_0 \text{ for all } i \in \mathbb{N}\}$$

of the simplex

$$\Delta = \{x = (x(1), x(2), \dots) : x(1) \geq x(2) \geq \dots \geq 0, |x|_1 \leq 1\},$$

where we write $|x|_p = (\sum_{i \in \mathbb{N}} |x(i)|^p)^{1/p}$.

First, we sample a Ξ^N -iid sequence $(x_k^N, k \in \mathbb{N})$ in Δ^N . Then, conditionally given $(x_k^N, k \in \mathbb{N})$, let $(\pi_k^N, k \in \mathbb{N})$ be a sequence of independent random partitions of $[N]$ such that for each $k \in \mathbb{N}$, the partition π_k^N is uniformly distributed on the set of partitions of $[N]$ whose block sizes are given by (any reordering of) Nx_k^N . In each generation $k \in \mathbb{N}$ of the population model, we partition the individuals into families, saying that individuals are in the same family if their labels are in the same block of π_k^N . Each family draws its common ancestor in generation $k-1$ uniformly without replacement. Tracing back the ancestral lineage, we denote by $A_j(k, i)$ the label of the ancestor in generation j of the individual i of generation k , for $j \in \mathbb{N}_0$ with $j \leq k$.

We are interested in the genealogical distances between the individuals in each generation. Given a distance matrix $\rho_0^N \in \mathcal{U}_N$, we define $\rho_0^N(i, j)$ as the genealogical distance between the individuals i and j in generation 0, for $i, j \in [N]$. Then we define the genealogical distance $\rho_\ell^N(i, j)$ between individuals i and j in a later generation $\ell \in \mathbb{N}$ by

$$\rho_\ell^N(i, j) = \begin{cases} 2c_N(\ell - \max\{k = 0, \dots, \ell : A_k(\ell, i) = A_k(\ell, j)\}) & \text{if } A_0(\ell, i) = A_0(\ell, j) \\ 2c_N\ell + \rho_0^N(A_0(\ell, i), A_0(\ell, j)) & \text{else,} \end{cases}$$

where we choose the scaling factor

$$(3.1) \quad c_N = \int \sum_{i=1}^N x(i) \frac{Nx(i) - 1}{N - 1} \Xi^N(dx).$$

We always assume $c_N > 0$. Analogously to e.g. [13, 16], the genealogical distance $\rho_\ell^N(i, j)$ between the individuals i and j in generation ℓ is, up to the scaling factor, the number of generations backwards until these individuals have the same ancestor if they have the same ancestor in generation 0, else $\rho_\ell^N(i, j)$ is given by the genealogical distance of their ancestors in generation 0. The quantity c_N is known as the pairwise coalescence probability, i.e. the probability that two individuals that are sampled uniformly without replacement from some generation $k \in \mathbb{N}$ have the same ancestor in generation $k-1$. Indeed, conditionally given x_k^N , the first sample is in any family with probability $x_k^N(i)$ when $Nx_k^N(i)$ is the size of that family. Conditionally given x_k^N and the first sample, the second sample is in the same family with probability $(Nx_k^N(i) - 1)/(N - 1)$.

In the next subsections, we state four invariance principles for processes that we read off from this population model. The invariance principles define the limit processes, but we recall the limit processes independently in Section 4. All our limit processes are characterized by the probability measures on the simplex Δ , we denote the space of these measures by $\mathcal{M}_1(\Delta)$. If $\Xi \in \mathcal{M}_1(\Delta)$ satisfies the condition

$$(3.2) \quad \Xi\{0\} > 0 \text{ or } \int |x|_1 |x|_2^{-2} \Xi(dx) = \infty,$$

then we speak of the dust-free case and we write $\Xi \in \mathcal{M}_{\text{nd}}$. The converse case is called the case with dust. We set $\mathcal{M}_{\text{dust}} = \mathcal{M}_1(\Delta) \setminus \mathcal{M}_{\text{nd}}$. For $\Xi \in \mathcal{M}_1(\Delta)$, we denote by Ξ_0 the measure on Δ with

$$(3.3) \quad \Xi = \Xi_0 + \Xi\{0\}\delta_0.$$

3.2. Processes with values in the space of metric measure spaces. Let $\chi_0^N \in \mathbb{U}_N$, and let ρ_0^N be a random variable with distribution ν^{N, χ_0^N} that is independent of the sequence $(\pi_k^N, k \in \mathbb{N})$ from Subsection 3.1. We define the process $(\rho_k^N, k \in \mathbb{N}_0)$ as in Subsection 3.1 from $(\pi_k^N, k \in \mathbb{N})$ and the initial state ρ_0^N . For $k \in \mathbb{N}_0$, we set

$$(3.4) \quad \chi_k^N = \llbracket [N], \rho_k^N, N^{-1} \sum_{i=1}^N \delta_i \rrbracket = \psi_N(\rho_k^N).$$

While ρ_k^N describes the genealogy of generation k as a leaf-labeled tree, the unlabeled tree is given by χ_k^N . We call the process $(\chi_k^N, k \in \mathbb{N}_0)$ a \mathbb{U}_N -valued Ξ^N -Cannings chain with initial state χ_0^N . We call this process also a tree-valued Ξ^N -Cannings chain.

Remark 3.1 (Markov property and transition kernel). We denote by \mathcal{P}_N the set of partitions of $[N]$. For $\pi \in \mathcal{P}_N$ and $i \in [N]$, we denote by $\pi(i)$ the integer k such that the k -th block of π contains i when the blocks are ordered increasingly according to their respective smallest element. As in [16], we associate with each element π of \mathcal{P}_N a transformation $\mathfrak{U}_N \rightarrow \mathfrak{U}_N$, which we also denote by π , by

$$(3.5) \quad \pi(\rho) = (\rho(\pi(i), \pi(j)))_{i,j \in [n]}.$$

We write $\underline{\underline{2}}_N = (\mathbf{1}\{i \neq j\})_{i,j \in [N]}$. The Markov property of $(\chi_k^N, k \in \mathbb{N}_0)$ follows as for each $k \in \mathbb{N}_0$, the distance matrix $\rho_{k+1}^N - c_N \underline{\underline{2}}_N$ has conditional distribution $\pi_{k+1}^N(\nu^{N, \chi_k^N})$ given π_{k+1}^N and $(\rho_j^N, j \leq k)$ by the construction in Section 3.1. We denote by p_N the transition kernel of $(\chi_k^N, k \in \mathbb{N}_0)$ which can be stated as

$$p_N(\chi, B) = \sum_{\pi \in \mathcal{P}_N} \mathbb{P}(\pi_1^N = \pi) \int \nu^{N, \chi}(d\rho) \mathbf{1}\left\{ \psi_N(\pi(\rho) + c_N \underline{\underline{2}}_N) \in B \right\}$$

for all $\chi \in \mathbb{U}_N$ and measurable $B \subset \mathbb{U}$. This transition kernel generalizes the one of a tree-valued Moran model from [13] or a tree-valued Λ -Cannings process that is discussed in [18].

For $c \in \mathbb{R}_+$, we set $\Delta_c = \{x \in \Delta : x(1) > c\}$.

Theorem 3.2. *Let $\Xi \in \mathcal{M}_{\text{nd}}$. Assume that χ_0^N converges to some χ_0 in $(\mathbb{U}, d_{\text{GP}})$ as N tends to infinity. Furthermore, assume that*

$$(3.6) \quad \lim_{N \rightarrow \infty} c_N = 0$$

and that

$$(3.7) \quad \begin{aligned} &\text{for arbitrarily small } \varepsilon > 0, \text{ on } \Delta_\varepsilon, \text{ the finite measures } c_N^{-1} \Xi^N(dx) \\ &\text{converge weakly to } |x|_2^{-2} \Xi(dx) \text{ as } N \text{ tends to infinity.} \end{aligned}$$

Then the processes $(\chi_{\lfloor c_N^{-1} t \rfloor}^N, t \in \mathbb{R}_+)$ converge in distribution to a \mathbb{U} -valued Ξ -Fleming-Viot process with initial state χ_0 in the space of càdlàg paths in $(\mathbb{U}, d_{\text{GP}})$, endowed with the Skorohod metric, as N tends to infinity.

Condition (3.6) is Condition (1.6) in [24] and ensures that the limit process is a process in continuous time with no fixed jumps. Condition (3.7) is Condition (2.9) of [24] and yields the convergence of the transition kernels to the generator of the limit process. The assumption $\Xi \in \mathcal{M}_{\text{nd}}$ in the theorem above is necessary as the \mathbb{U} -valued Ξ -Fleming-Viot process exists only for these Ξ . We include the case with dust in two ways: In Section 3.3, we consider the distance matrix distributions. In Section 3.4, we decompose the genealogical trees.

3.3. Processes of distance matrix distributions. We associate with the \mathbb{U}_N -valued Ξ^N -Cannings chain from the last subsection the process

$$(\xi_k^N, k \in \mathbb{N}_0) = (\nu^{\chi_k^N}, k \in \mathbb{N}_0)$$

with values in the space $(\mathcal{U}^{\text{erg}}, d_P)$ which we recalled in Section 2.2. Also $(\xi_k^N, k \in \mathbb{N}_0)$ is a Markov process. This follows from the Markov property of $(\chi_k^N, k \in \mathbb{N}_0)$ as ξ_k^N uniquely determines χ_k^N by the Gromov reconstruction theorem. As $(\xi_k^N, k \in \mathbb{N}_0)$ takes values in the space of distance matrix distributions

$$\mathcal{U}_N = \{\nu^\chi : \chi \in \mathbb{U}_N\} \subset \mathcal{U}^{\text{erg}},$$

we call this process a \mathcal{U}_N -valued Ξ^N -Cannings chain. When we do not want to specify the state space, we call also this process a tree-valued Ξ^N -Cannings chain.

Theorem 3.3. *Let $\Xi \in \mathcal{M}_1(\Delta)$. Assume that conditions (3.6) and (3.7) hold, and that ξ_0^N converges weakly to some probability measure ξ_0 on \mathfrak{U} as N tends to infinity. Then the processes $(\xi_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+)$ converge in distribution to a \mathcal{U}^{erg} -valued Ξ -Fleming-Viot process with initial state ξ_0 in the space of càdlàg paths in $(\mathcal{U}^{\text{erg}}, d_P)$, endowed with the Skorohod metric, as N tends to infinity.*

Remark 3.4. Consider the case that $\Xi \in \mathcal{M}_{\text{nd}}$ and that $\xi_0 = \nu^{\chi_0}$ for some $\chi_0 \in \mathbb{U}$. Let $(\chi_t, t \in \mathbb{R}_+)$ be a \mathbb{U} -valued Ξ -Fleming-Viot process with initial state χ_0 . Then, by the definition in [16, Section 8], the process $(\nu^{\chi_t}, t \in \mathbb{R}_+)$ is a \mathcal{U}^{erg} -valued Ξ -Fleming-Viot process and the assertion of Theorem 3.3 follows from Theorem 3.2 as the map $\mathbb{U} \rightarrow \mathcal{U}^{\text{erg}}, \chi \mapsto \nu^\chi$ is continuous.

3.4. Processes with values in the space of marked metric measure spaces. Recall the decomposition $\beta : \mathfrak{U}_N \rightarrow \hat{\mathfrak{U}}_N$ of ultrametric distance matrices from Section 2.1 and the construction $\hat{\psi}_N : \hat{\mathfrak{U}}_N \rightarrow \hat{\mathbb{U}}_N$ of marked metric measure spaces from marked distance matrices from Section 2.2. We define a process $(\hat{\chi}_k^N, k \in \mathbb{N})$ which we call a $\hat{\mathbb{U}}_N$ -valued Ξ^N -Cannings chain by

$$\hat{\chi}_k^N = \hat{\psi}_N \circ \beta(\rho_k^N)$$

for $k \in \mathbb{N}_0$, where $(\rho_k^N, k \in \mathbb{N}_0)$ is defined as in Section 3.2. More loosely, we also call the process $(\hat{\chi}_k^N, k \in \mathbb{N})$ a tree-valued Ξ^N -Cannings chain.

We denote by b_N the probability that an individual that is sampled uniformly from a fixed generation belongs to a family with more than one member. By construction,

$$b_N = \mathbb{E} \left[\sum_{i=1}^{\infty} x_1^N(i) \mathbf{1}_{\{x_1^N(i) > 1/N\}} \right].$$

Theorem 3.5. *Let $\Xi \in \mathcal{M}_1(\Delta)$. Assume that conditions (3.6) and (3.7) hold, and that $\hat{\chi}_0^N$ converges to some $\hat{\chi}_0$ in $(\hat{\mathbb{U}}, d_{\text{mGP}})$ as N tends to infinity.*

(i) *If $\Xi \in \mathcal{M}_{\text{dust}}$, assume in addition that*

$$(3.8) \quad \lim_{N \rightarrow \infty} b_N / c_N = \int |x|_1 |x|_2^{-2} \Xi(dx).$$

(ii) *If $\Xi \in \mathcal{M}_{\text{nd}}$, assume in addition that the marked metric measure space $\hat{\chi}_0$ supports only the zero mark.*

Then the processes $(\hat{\chi}_{[c_N^{-1}t]}^N, t \in \mathbb{R}_+)$ converge in distribution to a $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process with initial state $\hat{\chi}_0$ in the space of càdlàg paths in $(\hat{\mathbb{U}}, d_{\text{mGP}})$, endowed with the Skorohod metric, as N tends to infinity.

In the dust-free case, assumption (ii) is necessary for right continuity of the limit process at time 0. The expression on the right-hand side of (3.8) is the rate $\lambda_{1, \{\{1\}\}}$ in the martingale problem for the $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process which we recall in Section 4.2. In the lookdown construction from [16], $\lambda_{1, \{\{1\}\}}$ is the rate of reproduction events in which the individual on a fixed level belongs to a non-singleton family, which we will use in Remark 3.9. The expression on the right-hand side of (3.8) also occurs as the parameter of the limiting exponential distribution of the length of an external branch that is drawn randomly from a Ξ -coalescent as the sample size tends to infinity, see [21].

We also decompose some metric measure spaces analogously to Section 2.1: We denote also by β the function $\mathbb{U}_N \rightarrow \hat{\mathbb{U}}_N$ that maps χ to $\psi_N \circ \beta(\rho)$ where ρ is any element of \mathfrak{U}_N with $\psi_N(\rho) = \chi$. The map $\beta : \mathbb{U}_N \rightarrow \hat{\mathbb{U}}_N$ decomposes the (unlabeled) tree given by (the isomorphism class of) a metric measure space χ such that in $\beta(\chi)$, the lengths of the external branches are encoded by the marks, and the distances between their starting points are given by the metric. Note that $\hat{\chi}_k^N = \beta(\chi_k^N)$ for all $k \in \mathbb{N}_0$. The Markov property of the process $(\hat{\chi}_k^N, k \in \mathbb{N}_0)$ now follows from the Markov property of $(\chi_k^N, k \in \mathbb{N}_0)$ as χ_k^N is uniquely determined by $\beta(\chi_k^N)$.

3.5. Another decomposition of the evolving genealogical trees. In Theorem 3.5, the external branches of genealogical trees play a crucial role. However, for the proof in the case with dust, it seems more convenient to work with a different decomposition of the genealogical trees that does not bring about the freeing phenomenon which we mention below. We will therefore use Proposition 3.8 below in the proof of Theorem 3.5 in the case with dust. The process of these differently decomposed trees corresponds to the construction in [16, Section 7] where an infinite population is considered.

First we define a process $((r_k^N, v_k^N), k \in \mathbb{N}_0)$ of marked distance matrices. Let $\tilde{\chi}_0^N \in \hat{\mathbb{U}}_N$, let (r_0^N, v_0^N) be a $\hat{\mathfrak{U}}_N$ -valued random variable with distribution $\nu^{N, \tilde{\chi}_0^N}$, and let $\rho_0^N = \alpha(r_0^N, v_0^N)$, with α defined in Section 2.1. We define the process $(\rho_k^N, k \in \mathbb{N})$ as in Subsection 3.1 from the initial state ρ_0 and the sequence $(\pi_k^N, k \in \mathbb{N})$, assuming that $(\pi_k^N, k \in \mathbb{N})$ is independent of (r_0^N, v_0^N) .

For $\ell \in \mathbb{N}$, we define (r_ℓ^N, v_ℓ^N) as follows. For $i \in [N]$, if there exists a latest generation $k \in [\ell]$ in which the individual $A_k(\ell, i)$ is in a non-singleton block of π_k^N , we set $v_\ell^N(i) = c_N(\ell - k + 1)$. Else, that is, if $A_k(\ell, i)$ forms a singleton block of π_k^N for each $k \in [\ell]$, then we set $v_\ell^N(i) = c_N \ell + v_0^N(A_0(\ell, i))$. For $i, j \in [N]$, we set

$$r_\ell^N(i, j) = (\rho_\ell^N(i, j) - v_\ell^N(i) - v_\ell^N(j)) \mathbf{1}\{i \neq j\}.$$

Lemma 3.6. *It holds $(r_k^N, v_k^N) \in \hat{\mathfrak{U}}_N$ for all $k \in \mathbb{N}_0$.*

Proof. For $k \in \mathbb{N}$, we denote by $\cup \sigma_k^N$ the union of the non-singleton blocks of π_k^N (anticipating the notation of Remark 3.7 below). By construction,

$$\begin{aligned} r_k^N(i, j) = & v_{k-1}^N(A_{k-1}(k, i)) \mathbf{1}\{i \in \cup \sigma_k^N\} + r_{k-1}^N(A_{k-1}(k, i), A_{k-1}(k, j)) \\ & + v_{k-1}^N(A_{k-1}(k, j)) \mathbf{1}\{j \in \cup \sigma_k^N\} \end{aligned}$$

for all $k \in \mathbb{N}$ and all distinct $i, j \in [N]$. Under the assumption that r_{k-1}^N satisfies the triangle inequality, it is easily checked that also r_k^N satisfies the triangle inequality. The assertion follows by induction. \square

Analogously to [16, Remark 7.1], the quantity $v_k^N(i)$ can be interpreted as the age of the individual i of generation k . The quantity $v_k^N(i)$ needs not coincide with the (scaled) length of the external branch that ends in individual i in the (scaled) genealogical tree of generation k . Indeed, if $v_k(i) < k$ and no other members of the family of $A_{k-v_k^N(i)}(k, i)$ in generation

$k - v_k^N(i)$ have descendants in generation k , then this external branch is longer than $v_k^N(i)$. A related phenomenon in evolving coalescents is the so-called freeing where internal branches become part of external branches (see Dahmer and Kersting [7], in particular Figure 2 therein). The quantity $r_k^N(i, j)$ is the distance between the individuals that correspond to the parents in [15].

We obtain a $\hat{\mathbb{U}}_N$ -valued process $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$ by setting $\tilde{\chi}_k^N = \hat{\psi}_N(r_k^N, v_k^N)$ for $k \in \mathbb{N}$. The definition of (r_0^N, v_0^N) yields $\tilde{\chi}_0^N = \hat{\psi}_N(r_0^N, v_0^N)$.

Remark 3.7 (Markov property and transition kernel). We denote by \mathcal{S}_N the set of semi-partitions of $[N]$. A semi-partition σ of $[N]$ is a system of nonempty disjoint subsets of $[N]$ which we call blocks. We denote by $\cup \sigma$ the union of the blocks of σ (which can be a proper subset of $[N]$). For $\sigma \in \mathcal{S}_N$ and $i \in [N]$, we define $\sigma(i) = \pi(i)$ where π is the partition of $[n]$ that has the same non-singleton blocks as σ , that is, $\{B \in \sigma : \#B \geq 2\} = \{B \in \pi : \#B \geq 2\}$, and $\pi(i)$ is defined as in Remark 3.1. As in [16], we associate with each element σ of \mathcal{S}_N a transformation $\hat{\mathbb{U}}_N \rightarrow \hat{\mathbb{U}}_N$, which we also denote by σ , by $\sigma(r, v) = (r', v')$, where

$$v'(i) = v(\sigma(i)) \mathbf{1}\{i \notin \cup \sigma\}$$

and

$$r'(i, j) = (v(\sigma(i)) \mathbf{1}\{i \in \cup \sigma\} + r(\sigma(i), \sigma(j)) + v(\sigma(j)) \mathbf{1}\{j \in \cup \sigma\}) \mathbf{1}\{i \neq j\}$$

for $i, j \in [N]$.

For $k \in \mathbb{N}$, let σ_k^N be the semi-partition that consists of the non-singleton blocks of the partition π_k^N from Section 3.1, $\sigma_k^N = \{B \in \pi_k^N : \#B \geq 2\}$. We write $\mathbf{1}_N = (1)_{i \in [N]}$. The Markov property of $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$ follows as by construction, for each $k \in \mathbb{N}_0$, the marked distance matrix $(r_{k+1}^N, v_{k+1}^N - c_N \mathbf{1}_N)$ has conditional distribution $\sigma_{k+1}^N(\nu^{N, \tilde{\chi}_k^N})$ given σ_{k+1}^N and $((r_j^N, v_j^N), j \leq k)$. We denote by \tilde{p}_N the transition kernel of $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$ which can be stated as

$$\tilde{p}_N(\chi, B) = \sum_{\sigma \in \mathcal{S}_N} \mathbb{P}(\sigma_1^N = \sigma) \int \nu^{N, \chi}(d(r, v)) \mathbf{1}\left\{\hat{\psi}_N(\sigma(r, v) + c_N(0, \mathbf{1}_N)) \in B\right\}$$

for all $\chi \in \hat{\mathbb{U}}_N$ and measurable $B \subset \hat{\mathbb{U}}$.

Proposition 3.8. *Assume that assumptions of Theorem 3.5 hold and that $\tilde{\chi}_0^N$ converges to $\hat{\chi}_0$ in $(\hat{\mathbb{U}}, d_{\text{mGP}})$ as N tends to infinity. Then the processes $(\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+)$ converge in distribution to a $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process with initial state $\hat{\chi}_0$ in the space of càdlàg paths in $(\hat{\mathbb{U}}, d_{\text{mGP}})$, endowed with the Skorohod metric, as N tends to infinity.*

Remark 3.9. In Proposition 3.8, the assumption (3.8) is necessary in case $\Xi \in \mathcal{M}_{\text{dust}}$. To see this, let $t \in (0, \infty)$, and let $v_t = (v_t(i), i \in \mathbb{N})$ be defined as in [16, Section 7.2]. By the lookdown construction in [16], the truncated first entry $v_t(1) \wedge t$ is equal in distribution to $T \wedge t$ for an exponentially distributed random variable T with parameter $\int |x|_1 |x|_2^{-2} \Xi_0(dx)$. Moreover, by the construction in Subsection 3.1, the random variable $v_{\lfloor c_N^{-1}t \rfloor}^N(1) \wedge (c_N \lfloor c_N^{-1}t \rfloor)$ is distributed as $c_N(T_N \wedge \lfloor c_N^{-1}t \rfloor)$ for a geometrically distributed random variable T_N with parameter b_N . Let $(\hat{\chi}_s, s \in \mathbb{R}_+)$ be a $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process with initial state $\hat{\chi}_0$ (as defined in [15, 16]) and let $\varpi : \hat{\mathbb{U}} \rightarrow \mathbb{R}_+$, $(r, v) \mapsto v(1)$. Then, as in [16, Proposition 4.7], the random variable $v_t(1)$ has distribution $\mathbb{E}[\varpi(\nu^{\hat{\chi}_t})]$. Moreover, as $v_{\lfloor c_N^{-1}t \rfloor}^N$ is exchangeable, the random variable $v_{\lfloor c_N^{-1}t \rfloor}^N(1)$ has distribution $\mathbb{E}[\varpi(\nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})]$.

Now assume that $(\tilde{\chi}_{\lfloor c_N^{-1}s \rfloor}^N, s \in \mathbb{R}_+)$ converges as asserted in Proposition 3.8. Then, as the path $(\hat{\chi}_s, s \in \mathbb{R}_+)$ is a.s. continuous at fixed times, also $\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N$ converges in distribution to $\hat{\chi}_t$. By continuity of the map ϖ , also $v_{\lfloor c_N^{-1}t \rfloor}^N(1)$ converges in distribution to $v_t(1)$. As t can be chosen arbitrarily large, it follows that $c_N T_N$ converges in distribution to T which implies condition (3.8).

4. TREE-VALUED FLEMING-VIOT PROCESSES

In this section, we recall from [16] the martingale problems for tree-valued Ξ -Fleming-Viot processes. Path regularity is shown in [15]. (In [15, 16], tree-valued Ξ -Fleming-Viot processes are considered for finite measures Ξ on Δ .) For $n \in \mathbb{N}$, we define the restrictions

$$\gamma_n : \mathfrak{U} \cup \bigcup_{\ell \geq n} \mathfrak{U}_\ell \rightarrow \mathfrak{U}_n, \quad \rho \mapsto (\rho(i, j))_{i, j \in [n]}$$

and

$$\gamma_n : \hat{\mathfrak{U}} \cup \bigcup_{\ell \geq n} \hat{\mathfrak{U}}_\ell \rightarrow \hat{\mathfrak{U}}_n, \quad (r, v) \mapsto ((r(i, j))_{i, j \in [n]}, (v(i))_{i \in [n]}).$$

Let \mathcal{C}_n denote the set of bounded differentiable functions $\mathbb{R}^{n^2} \rightarrow \mathbb{R}$ with bounded uniformly continuous derivative. For $\phi \in \mathcal{C}_n$, we write also ϕ for the function $\phi \circ \gamma_n$. For $\phi \in \mathcal{C}_n$, we call the function $\mathbb{U} \rightarrow \mathbb{R}$, $\chi \mapsto \nu^\chi \phi$ the polynomial associated with ϕ . As in [19, Corollary 2.8], the algebra of polynomials

$$\Pi = \{\mathbb{U} \rightarrow \mathbb{R}, \chi \mapsto \nu^\chi \phi : n \in \mathbb{N}, \phi \in \mathcal{C}_n\}$$

is convergence determining in \mathbb{U} .

Analogously, let $\hat{\mathcal{C}}_n$ be the set of bounded differentiable functions $\phi : \mathbb{R}^{n^2} \times \mathbb{R}^n \rightarrow \mathbb{R}$ with bounded uniformly continuous derivative. For $\phi \in \hat{\mathcal{C}}_n$, we denote also the function $\phi \circ \gamma_n$ by ϕ , and we associate with ϕ the marked polynomial $\hat{\mathbb{U}} \rightarrow \mathbb{R}$, $\chi \mapsto \nu^\chi \phi$. The algebra of marked polynomials

$$\hat{\Pi} = \{\hat{\mathbb{U}} \rightarrow \mathbb{R}, \chi \mapsto \nu^\chi \phi : n \in \mathbb{N}, \phi \in \hat{\mathcal{C}}_n\}$$

is convergence determining in $\hat{\mathbb{U}}$, see [19, Corollary 2.8]. The algebra

$$\mathcal{C} = \{\mathcal{U}^{\text{erg}} \rightarrow \mathbb{R}, \nu \mapsto \nu \phi : n \in \mathbb{N}, \phi \in \mathcal{C}_n\}$$

is convergence determining in \mathcal{U}^{erg} , see [16, Remark 4.5].

4.1. The \mathbb{U} -valued Ξ -Fleming-Viot process. A \mathbb{U} -valued Ξ -Fleming-Viot process exists for $\Xi \in \mathcal{M}_{\text{nd}}$ and any initial state in \mathbb{U} . It is a Markov process and has a version with càdlàg paths and no discontinuities at fixed times. This process is the unique solution of the martingale problem (B, Π) , we recall the generator B in this subsection. For $n \in \mathbb{N}$, let \mathcal{P}_n be the set of partitions of $[n]$, and let κ_n^∞ be the probability kernel from Δ to \mathcal{P}_n given by Kingman's correspondence, which we also recall here as we need it in Section 6.2. Consider independent uniform $[0, 1]$ -valued random variables U_1, \dots, U_n . For $x \in \Delta$, let $\kappa_n^\infty(x, \cdot)$ be the distribution of the random partition in \mathcal{P}_n such that two different integers i and j are in the same block if and only if there exists $\ell \in \mathbb{N}$ with

$$U_i, U_j \in \left(\sum_{k=1}^{\ell-1} x(k), \sum_{k=1}^{\ell} x(k) \right).$$

For $\phi \in \mathcal{C}_n$, we define the function

$$\langle \nabla \phi, \underline{2} \rangle : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}, \quad \rho \mapsto 2 \sum_{\substack{i, j \in \mathbb{N} \\ i \neq j}} \frac{\partial}{\partial \rho(i, j)} \phi(\rho).$$

Let $\mathbf{0}_n = \{\{1\}, \dots, \{n\}\}$, and recall Ξ_0 from equation (3.3). For $\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}$, we define

$$\begin{aligned} \lambda_\pi &= \int \kappa_n^\infty(x, \pi) |x|_2^{-2} \Xi_0(dx) \\ &\quad + \Xi\{\mathbf{0}\} \mathbf{1}\{\pi \text{ contains one doubleton and apart from that only singletons}\}. \end{aligned}$$

The rates λ_π are equal to those of Schweinsberg [25], see [16, Remark 6.1]. Moreover, we associate with each element of \mathcal{P}_n a transformation on \mathfrak{U}_n as in Remark 3.1.

Now we set $B = B_{\text{grow}} + B_{\text{res}}$ with

$$B_{\text{grow}}\Phi(\chi) = \int \nu^\chi(d\rho) \langle \nabla \phi, \underline{\underline{2}} \rangle(\rho)$$

and

$$B_{\text{res}}\Phi(\chi) = \sum_{\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}} \lambda_\pi \int \nu^\chi(d\rho) (\phi(\pi(\gamma_n(\rho))) - \phi(\rho))$$

for $\chi \in \mathbb{U}$, $n \in \mathbb{N}$, and $\phi \in \mathcal{C}_n$ with associated polynomial Φ .

4.2. The $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process. We first consider the case $\Xi \in \mathcal{M}_{\text{nd}}$. Using the isometry

$$\beta_0 : \mathbb{U} \rightarrow \hat{\mathbb{U}}, \quad \llbracket X, \rho, \mu \rrbracket \mapsto \llbracket X, \rho, \mu \otimes \delta_0 \rrbracket$$

which adds a mark component that is concentrated in zero, we define for $\chi \in \mathbb{U}$ a $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process with initial state $\beta_0(\chi)$ by $(\beta_0(\chi_t), t \in \mathbb{R}_+)$, where $(\chi_t, t \in \mathbb{R}_+)$ is a \mathbb{U} -valued Ξ -Fleming-Viot process with initial state χ and càdlàg paths. Also this process is Markov and has càdlàg paths with no discontinuities at fixed times.

Now we consider the case $\Xi \in \mathcal{M}_{\text{dust}}$. Then for each initial state in $\hat{\mathbb{U}}$, there exists a $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process which is a Markov process and has a version with càdlàg paths and no discontinuities at fixed times. This process is the unique solution of the martingale problem $(\hat{B}, \hat{\Pi})$, we recall the generator \hat{B} now. For $n \in \mathbb{N}$, we define the set \mathcal{S}_n of semi-partitions of $[n]$ and the transformation on $\hat{\mathbb{U}}_n$ associated with each element of \mathcal{S}_n as in Remark 3.7.

Let K_n^∞ be the probability kernel from Δ to \mathcal{S}_n defined as follows: Consider independent uniform $[0, 1]$ -valued random variables U_1, \dots, U_n . For $x \in \Delta$, let $K_n^\infty(x, \cdot)$ be the distribution of the random element σ of \mathcal{S}_n such that any two integers i and j are in a common subset $B \in \sigma$ if and only if there exists $\ell \in \mathbb{N}$ with

$$U_i, U_j \in \left(\sum_{k=1}^{\ell-1} x(k), \sum_{k=1}^{\ell} x(k) \right).$$

For $\phi \in \hat{\mathcal{C}}_n$, we define the function

$$\langle \nabla^v \phi, \underline{\underline{1}} \rangle : \mathbb{R}^{\mathbb{N}^2} \times \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}, \quad (r, v) \mapsto \sum_{i \in \mathbb{N}} \frac{\partial}{\partial v(i)} \phi(r, v).$$

For $\sigma \in \mathcal{S}_n \setminus \{\emptyset\}$, we define the rates

$$\lambda_{n,\sigma} = \int K_n^\infty(x, \sigma) |x|_2^{-2} \Xi_0(dx).$$

These rates are finite by the assumption that $\Xi \in \mathcal{M}_{\text{dust}}$ and as $K_n^\infty(x, \sigma) \leq K_1^\infty(x, \{\{1\}\}) = |x|_1$ for all $x \in \Delta$. Now we set

$$\hat{B} = \hat{B}_{\text{grow}} + \hat{B}_{\text{res}},$$

$$\hat{B}_{\text{grow}}\Phi(\chi) = \int \nu^\chi(dr dv) \langle \nabla^v \phi, \underline{\underline{1}} \rangle(r, v)$$

and

$$\hat{B}_{\text{res}}\Phi(\chi) = \sum_{\sigma \in \mathcal{S}_n \setminus \{\emptyset\}} \lambda_{n,\sigma} \int \nu^\chi(dr dv) (\phi(\sigma(\gamma_n(r, v))) - \phi(r, v))$$

for $\chi \in \hat{\mathbb{U}}$, $n \in \mathbb{N}$, and $\phi \in \hat{\mathcal{C}}_n$ with associated marked polynomial Φ .

4.3. The \mathcal{U}^{erg} -valued Ξ -Fleming-Viot process. A \mathcal{U}^{erg} -valued Ξ -Fleming-Viot process exists for every $\Xi \in \mathcal{M}_1(\Delta)$ and every initial state in \mathcal{U}^{erg} . It is a Markov process and has a version with càdlàg paths and no discontinuities at fixed times. This process is the unique solution of the martingale problem (C, \mathcal{C}) which is defined as follows. Let the rates λ_π for $\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}$, $n \in \mathbb{N}$ be defined from the measure $\Xi \in \mathcal{M}_1(\Delta)$ as in Subsection 4.1. We define the generator C by

$$\begin{aligned} C &= C_{\text{grow}} + C_{\text{res}} \\ C_{\text{grow}}\Psi(\xi) &= \int \xi(d\rho) \langle \nabla \phi, \underline{2} \rangle(\rho) \\ C_{\text{res}}\Psi(\xi) &= \sum_{\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}} \lambda_\pi \int \xi(d\rho) (\phi(\pi(\gamma_n(\rho))) - \phi(\rho)) \end{aligned}$$

for $\xi \in \mathcal{U}^{\text{erg}}$, $n \in \mathbb{N}$, $\phi \in \mathcal{C}_n$, and $\Psi \in \mathcal{C}$, $\Psi : \xi' \mapsto \xi' \phi$.

If $(\hat{\chi}_t, t \in \mathbb{R}_+)$ is a $\hat{\mathcal{U}}$ -valued Ξ -Fleming-Viot process, then the process $(\alpha(\nu^{\hat{\chi}_t}), t \in \mathbb{R}_+)$ of the associated distance matrix distributions is a \mathcal{U}^{erg} -valued Ξ -Fleming-Viot process.

5. AN EXAMPLE

In this section, we consider a sequence of Cannings models that does not satisfy the assertion of Theorem 3.5, but all its assumptions for $\Xi \in \mathcal{M}_{\text{dust}}$ except condition (3.8). The convergence of the $\hat{\mathcal{U}}_N$ -valued Cannings chains to the $\hat{\mathcal{U}}$ -valued Ξ -Fleming-Viot process in Theorem 3.5 is excluded as the length of a randomly sampled external branch converges to zero in distribution, while this quantity is a.s. non-zero at a fixed time for a $\hat{\mathcal{U}}$ -valued Fleming-Viot process in the case with dust. The convergence in Theorem 3.3 nevertheless holds. Here it comes to bear that the length of the first external branch is not a continuous functional of the infinite ultrametric that describes the genealogy. In the example of the present section, the length of a randomly sampled external branch converges to zero due to “perturbative” reproduction events that occur at high rate. However, these reproduction events are not visible in the limiting genealogy as each of them affects only a small part of the population.

We now work in the context of Section 3.1. We assume in this section that for N sufficiently large, the law Ξ^N of the sequence of the family sizes $x = (x(i), i \in \mathbb{N})$ in a reproduction event is given by the outcome of a two-step random experiment that can be described as follows. In the first step, we draw the type of the reproduction event from the set {ordinary, perturbative, trivial}. With probability N^{-1} , a reproduction event shall be ordinary, with probability $N^{-1/2}$ perturbative, and with probability $1 - N^{-1} - N^{-1/2}$ trivial. In the second step, the sequence x of the family sizes is drawn depending on the type of the reproduction event:

- (i) If the reproduction event is ordinary, sample the largest family size $Nx(1)$ according to the binomial distribution with number of trials N and success probability $1/2$. That is, each individual belongs to the largest family independently with probability $1/2$. Let the other families be singletons, so that $Nx(i) \in \{0, 1\}$ for $i \geq 2$.
- (ii) If the reproduction event is perturbative, sample the largest family size $Nx(1)$ according to the binomial distribution with parameters N and $N^{-1/3}$. Let the other families be singletons, so that $Nx(i) \in \{0, 1\}$ for $i \geq 2$.
- (iii) If the reproduction event is trivial, let all families be singletons, so that $Nx(i) = 1$ for $i \in [N]$, and $Nx(i) = 0$ for $i > N$.

The pairwise coalescence probability for this reproduction law equals

$$c_N = N^{-1}(\tfrac{1}{2})^2 + N^{-1/2}(N^{-1/3})^2$$

and is asymptotically equivalent to $\frac{1}{4}N^{-1}$ as N tends to infinity. The probability b_N is bounded from below by

$$N^{-1/2}N^{-1/3}(1 - N^{-(1/3) \cdot (N-1)}).$$

This is the probability that a perturbative reproduction event occurs in which the individual labeled by 1 belongs to a non-singleton family. Condition (3.7) is satisfied with $\Xi = \delta_{(1/2,0,0,\dots)} \in \mathcal{M}_{\text{dust}}$. As b_N/c_N tends to infinity as N tends to infinity, condition (3.8) is not satisfied.

In this section, we fix $t \in (0, \infty)$.

Proposition 5.1. *For $N \in \mathbb{N}$, let $(\hat{\chi}_k^N, k \in \mathbb{N}_0)$ be a $\hat{\mathbb{U}}$ -valued Cannings chain defined as in Section 3.4 from the measure Ξ^N . Let $(\hat{\chi}_s, s \in \mathbb{R}_+)$ be a $\hat{\mathbb{U}}$ -valued $\delta_{(1/2,0,0,\dots)}$ -Fleming-Viot process. Then $\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N$ does not converge in distribution to $\hat{\chi}_t$ as N tends to infinity.*

Remark 5.2. Proposition 5.1 implies that the processes $(\hat{\chi}_{\lfloor c_N^{-1}s \rfloor}^N, s \in \mathbb{R}_+)$ do not converge in distribution in the Skorohod metric to $(\hat{\chi}_s, s \in \mathbb{R}_+)$ as N tends to infinity. This follows as $(\hat{\chi}_s, s \in \mathbb{R}_+)$ does a.s. not jump at fixed times.

To prove Proposition 5.1, we use the following lemma which states that for large N , in the leaf-labeled genealogical tree of generation $\lfloor c_N^{-1}t \rfloor$, the length of the external branch that ends in individual 1 is typically small.

Lemma 5.3. *Let $(\rho_k^N, k \in \mathbb{N}_0)$ be defined as in Section 3 from the measure Ξ^N . Let $\varepsilon \in (0, t)$ and*

$$v^N = (v^N(1), \dots, v^N(N)) = \Upsilon(\rho_{\lfloor c_N^{-1}t \rfloor}^N).$$

Then, $\mathbb{P}(v^N(1) > \varepsilon) < 4\varepsilon$ for sufficiently large N .

Proof of Proposition 5.1. Let $\varpi : \hat{\mathbb{U}} \rightarrow \mathbb{R}_+$, $(r, v) \mapsto v(1)$. By exchangeability, the first entry $v^N(1)$ of the vector v^N in Lemma 5.3 has distribution $\mathbb{E}[\varpi(\nu^{\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})]$. By Lemma 5.3, the random variable $v^N(1)$ converges to zero in probability.

To obtain a contradiction, we assume that $\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N$ converges in distribution to $\hat{\chi}_t$. Then as in Remark 3.9, also the probability distributions $\mathbb{E}[\varpi(\nu^{\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})]$ converge weakly to $\mathbb{E}[\varpi(\nu^{\hat{\chi}_t})]$. But as in Remark 3.9, the measure $\mathbb{E}[\varpi(\nu^{\hat{\chi}_t})]$ is not the Dirac measure in zero. \square

Proof of Lemma 5.3. Let A_N be the number of ancestors in generation $\lfloor c_N^{-1}t \rfloor - \lfloor c_N^{-1}\varepsilon \rfloor$ of the individuals of generation $\lfloor c_N^{-1}t \rfloor$. Then,

$$(5.1) \quad \mathbb{P}(v^N(1) > \varepsilon, A_N \geq N/2) \leq (1 - N^{-1/2}N^{-1/3}(1 - (1 - N^{-1/3})^{N/2-1}))^{\lfloor c_N^{-1}\varepsilon \rfloor}.$$

We sketch a proof of the bound (5.1). The number of ancestors in generation k of the individuals of generation $\lfloor c_N^{-1}t \rfloor$ is non-decreasing in k for $k \leq \lfloor c_N^{-1}t \rfloor$. On the event $\{v^N(1) > \varepsilon, A_N \geq N/2\}$, no reproduction event in which the individual labeled by $A_k(\lfloor c_N^{-1}t \rfloor, 1)$ and another one of these ancestors are in the same block occurs in any generation k with $\lfloor c_N^{-1}t \rfloor - \lfloor c_N^{-1}\varepsilon \rfloor < k \leq \lfloor c_N^{-1}t \rfloor$. The right-hand side of inequality (5.1) is the probability that in none of the generations k with $\lfloor c_N^{-1}t \rfloor - \lfloor c_N^{-1}\varepsilon \rfloor < k \leq \lfloor c_N^{-1}t \rfloor$, the reproduction event is perturbative and individual $A_k(\lfloor c_N^{-1}t \rfloor, 1)$ is in the same family as any other of the at least $N/2$ many ancestors of the individuals of generation $\lfloor c_N^{-1}t \rfloor$.

For N sufficiently large, $\lfloor c_N^{-1}\varepsilon \rfloor \geq 3\varepsilon N$. As

$$\log(1 - N^{-1/3})^{N/2-1} \leq -(N/2 - 1)N^{-1/3} \rightarrow -\infty \quad (N \rightarrow \infty),$$

the right hand side of inequality (5.1) is bounded from above by $(1 - N^{-5/6}/2)^{3\varepsilon N}$ for N sufficiently large. This expression converges to zero as N tends to infinity.

It now suffices to show $\limsup_{N \rightarrow \infty} \mathbb{P}(A_N < N/2) < 4\varepsilon$. The event that no ordinary reproduction events occur between generations $\lfloor c_N^{-1}t \rfloor - \lfloor c_N^{-1}\varepsilon \rfloor$ and $\lfloor c_N^{-1}t \rfloor$ has probability at least $(1 - N^{-1})^{4\varepsilon N}$. Note that on this event, the random variable $N - A_N$ is stochastically bounded from above by $\sum_{i=1}^X Y_i$ where X, Y_1, Y_2, \dots are independent random variables, X is binomially distributed with parameters $\lfloor 4\varepsilon N \rfloor$ and $N^{-1/2}$, and all Y_i are binomially distributed with parameters N and $N^{-1/3}$. Here X corresponds to the number of perturbative reproduction events,

and Y_i corresponds to the decrease in the number of ancestors, backwards in time, in the i -th of these reproduction events. Then we have

$$(5.2) \quad \mathbb{P}(A_N < N/2) \leq 1 - (1 - N^{-1})^{4\epsilon N} + \mathbb{P}\left(\sum_{i=1}^X Y_i > N/2\right).$$

We estimate

$$(5.3) \quad \mathbb{P}\left(\sum_{i=1}^X Y_i > N/2\right) \leq \mathbb{P}(X > 8\epsilon N^{1/2}) + 8\epsilon N^{1/2} \mathbb{P}(Y_1 > (8\epsilon N^{1/2})^{-1} N/2).$$

By the Chebychev inequality, the first summand is bounded from above by

$$(8\epsilon N^{1/2} - 4\epsilon N^{1/2})^{-2} 4\epsilon N^{1/2},$$

and the second summand is bounded from above by

$$8\epsilon N^{1/2} ((8\epsilon N^{1/2})^{-1} N/2 - N^{2/3})^{-2} N^{2/3}.$$

Hence, the expression on the right-hand side of (5.3) tends to zero, and the expression on the right-hand side of (5.2) converges to $1 - e^{-4\epsilon} < 4\epsilon$ as $N \rightarrow \infty$. \square

6. CONVERGENCE OF THE TRANSITION KERNELS

This section contains the proofs of Theorems 3.2 and 3.3, and the proof for the case $\Xi \in \mathcal{M}_{\text{nd}}$ in Proposition 3.8. We write γ_n also for the restriction maps $\mathcal{S}_N \rightarrow \mathcal{S}_n$ and $\mathcal{P}_N \rightarrow \mathcal{P}_n$ for $n \leq N$ (that is, $\gamma_n(\sigma) = \{B \cap [n] : B \in \sigma\} \setminus \{\emptyset\}$). We define the rates λ_π and $\lambda_{n,\sigma}$ from the measure Ξ as in Section 4. We will need the following lemmas.

Lemma 6.1. *Let $\Xi \in \mathcal{M}_1(\Delta)$ and assume that condition (3.7) holds. Then,*

$$\lim_{N \rightarrow \infty} c_N^{-1} \mathbb{P}(\gamma_n(\pi_1^N) = \pi) = \lambda_\pi$$

for all $n \in \mathbb{N}$ and $\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}$.

Lemma 6.2. *Let $\Xi \in \mathcal{M}_{\text{dust}}$. Assume that conditions (3.7) and (3.8) hold. Then,*

$$\lim_{N \rightarrow \infty} c_N^{-1} \mathbb{P}(\gamma_n(\sigma_1^N) = \sigma) = \lambda_{n,\sigma}$$

for all $n \in \mathbb{N}$ and $\sigma \in \mathcal{S}_n \setminus \{\emptyset\}$.

Remark 6.3. Note that the assumption of Lemma 6.1 is condition (2.9) in [24, Theorem 2.1], and that the limits in the assertion of Lemma 6.1 are the limits (16) in [22, Theorem 2.1]. In Subsection 6.2, we prove Lemmas 6.1 and 6.2 directly, using continuity properties in particular of the probability kernel associated with Kingman's correspondence.

6.1. Proofs of invariance principles.

Proof of Theorem 3.2. Let $n \in \mathbb{N}$, $N \geq n$, and $\phi \in \mathcal{C}_n$. As in [13], we associate with ϕ not only the polynomial $\Phi : \mathbb{U} \rightarrow \mathbb{R}$, but also the N -polynomial

$$\Phi_N : \mathbb{U}_N \rightarrow \mathbb{R}, \quad \chi \mapsto \nu^{N,\chi} \phi.$$

As in Remark 3.1, let \mathcal{P}_N denote the transition kernel of the Markov chain $(\chi_k^N, k \in \mathbb{N}_0)$. Then,

$$\begin{aligned} p_N \Phi_N(\chi_0^N) &= \mathbb{E}[\Phi_N(\chi_1^N)] = \mathbb{E}[\phi(\rho_1^N)] \\ &= \sum_{\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}} \mathbb{P}(\gamma_n(\pi_1^N) = \pi) \int \nu^{N,\chi_0^N}(d\rho) \phi(\pi(\gamma_n(\rho))) + c_N \underline{\mathbb{2}}_n. \end{aligned}$$

For the second equality, we use that ν^{N,χ_k^N} is the conditional distribution of ρ_1^N given χ_1^N which follows as ρ_1^N is exchangeable. By the construction in Section 3.1, the conditional distribution

of $\gamma_n(\rho_1^N) - c_N \underline{2}_n$ given $\gamma_n(\pi_1^N)$ equals $\gamma_n(\pi_1^N)(\gamma_n(\nu^{N, \chi_0^N}))$. This yields the third equality. Now we have

$$\begin{aligned}
& c_N^{-1}(p_N - I)\Phi_N(\chi) \\
&= \mathbb{P}(\gamma_n(\pi_1^N) = \mathbf{0}_n) \int \nu^{N, \chi}(d\rho) c_N^{-1}(\phi(\rho + c_N \underline{2}_N) - \phi(\rho)) \\
(6.1) \quad & + \sum_{\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}} c_N^{-1} \mathbb{P}(\gamma_n(\pi_1^N) = \pi) \int \nu^{N, \chi}(d\rho) (\phi(\pi(\gamma_n(\rho)) + c_N \underline{2}_n) - \phi(\rho))
\end{aligned}$$

for all $\chi \in \mathbb{U}_N$. As $c_N = \mathbb{P}(\gamma_2(\pi_1^N) = \{\{1, 2\}\})$, by exchangeability of π_1^N , and by assumption (3.6),

$$1 - \mathbb{P}(\gamma_n(\pi_1^N) = \mathbf{0}_n) \leq \binom{n}{2} c_N \rightarrow 0 \quad (N \rightarrow \infty).$$

As $\phi \in \mathcal{C}_n$,

$$\lim_{N \rightarrow \infty} \sup_{\rho \in \mathfrak{U}_N} \left| c_N^{-1}(\phi(\rho + c_N \underline{2}_N) - \phi(\rho)) - \langle \nabla \phi, \underline{2} \rangle(\rho) \right| = 0,$$

this follows from the mean value theorem and uniform continuity of the derivative. Furthermore, for every bounded measurable function $f : \mathbb{R}^{n^2} \rightarrow \mathbb{R}$,

$$(6.2) \quad \int \nu^{N, \chi}(d\rho) f(\gamma_n(\rho)) - \int \nu^\chi(d\rho) f(\gamma_n(\rho)) \leq 2 \sup |f| n^2 / N.$$

This follows as for every semi-metric measure space $\chi = \llbracket [N], \rho, N^{-1} \sum_{i=1}^N \delta_i \rrbracket \in \mathbb{U}_N$, we can couple $\gamma_n(\nu^{N, \chi})$ and $\gamma_n(\nu^\chi)$ by sampling n times from $[N]$ uniformly with replacement and accepting this as a sample without replacement on the event that no element of $[N]$ is drawn more than once. The probability of the complementary event is bounded from above by

$$1 - (1 - n/N)^n \leq n^2 / N.$$

Taking $f(\rho) = \langle \nabla \phi, \underline{2} \rangle(\rho)$, we obtain from the bound (6.2)

$$\begin{aligned}
& \left| \int \nu^{N, \chi}(d\rho) c_N^{-1}(\phi(\rho + c_N \underline{2}_N) - \phi(\rho)) - \int \nu^\chi(d\rho) \langle \nabla \phi, \underline{2} \rangle(\rho) \right| \\
& \leq \int \nu^{N, \chi}(d\rho) \left| c_N^{-1}(\phi(\rho + c_N \underline{2}_N) - \phi(\rho)) - \langle \nabla \phi, \underline{2} \rangle(\rho) \right| + 2 \sup |\langle \nabla \phi, \underline{2} \rangle| n^2 / N.
\end{aligned}$$

With $f(\rho) = \phi(\pi(\rho)) - \phi(\rho)$, we obtain from (6.2)

$$\begin{aligned}
& \left| \int \nu^{N, \chi}(d\rho) (\phi(\pi(\gamma_n(\rho)) + c_N \underline{2}_N) - \phi(\rho)) - \int \nu^\chi(d\rho) (\phi(\pi(\gamma_n(\rho))) - \phi(\rho)) \right| \\
& \leq \sup \{ |\phi(\rho') - \phi(\rho)| : \rho, \rho' \in \mathfrak{U}_n, \|\rho - \rho'\| \leq c_N \} \\
& \quad + 4 \sup |\phi| n^2 / N.
\end{aligned}$$

Using also Lemma 6.1, we now obtain from equation (6.1) and the definition of B in Section 4.1 the convergence

$$\lim_{N \rightarrow \infty} \sup_{\chi \in \mathbb{U}_N} |c_N^{-1}(p_N - I)\Phi_N(\chi) - B\Phi(\chi)| = 0.$$

The algebra Π of polynomials strongly separates points in \mathbb{U} by [3, Lemma 4] and as Π generates the topology on \mathbb{U} . Let L denote the closure of Π for the supremum norm in the space of bounded continuous functions on \mathbb{U} . Analogously to Corollaries 9.3 and 9.4 in [16], the closure of B generates the semigroup on L of a \mathbb{U} -valued Ξ -Fleming-Viot process, which is strongly continuous. The assertion now follows from Corollary 4.8.9 in [11], condition (h) therein is satisfied. To see that the limit on the left-hand side of equation (8.48) on p. 234 of [11] equals zero, we set $f = \phi$ in equation (6.2). \square

Proof of Theorem 3.3. Let $n \in \mathbb{N}$ and $N \geq n$. Let p'_N be the transition kernel of the Markov chain $(\xi_k^N, k \in \mathbb{N}_0)$. Let $\phi \in \mathcal{C}_n$,

$$\Psi : \mathcal{U}^{\text{erg}} \rightarrow \mathbb{R}, \quad \xi \mapsto \xi\phi,$$

and

$$\Psi_N : \mathcal{U}_N \rightarrow \mathbb{R}, \quad \nu^X \mapsto \nu^{N,X}\phi.$$

Then, as in the proof of Theorem 3.2,

$$p'_N \Psi_N(\xi_0^N) = \mathbb{E}[\Psi_N(\xi_1^N)] = \mathbb{E}[\nu^{N,\chi_1^N} \phi] = \mathbb{E}[\phi(\rho_1^N)]$$

and

$$\begin{aligned} & c_N^{-1}(p_N - I)\Psi_N(\nu^X) \\ &= \mathbb{P}(\gamma_n(\pi_1^N) = \mathbf{0}_n) \int \nu^{N,X}(d\rho) c_N^{-1}(\phi(\rho + c_N \underline{\mathbf{2}}_N) - \phi(\rho)) \\ &+ \sum_{\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}} c_N^{-1} \mathbb{P}(\gamma_n(\pi_1^N) = \pi) \int \nu^{N,X}(d\rho) (\phi(\pi(\gamma_n(\rho)) + c_N \underline{\mathbf{2}}_n) - \phi(\rho)) \end{aligned}$$

for all $\chi \in \mathbb{U}_N$. As in the proof of Theorem 3.2, we have

$$\lim_{N \rightarrow \infty} \sup_{\chi \in \mathbb{U}_N} |c_N^{-1}(p_N - I)\Psi_N(\nu^X) - C\Psi(\nu^X)| = 0.$$

The algebra \mathcal{C} strongly separates points in \mathcal{U}^{erg} by [3, Lemma 4] and as \mathcal{C} generates the weak topology on \mathcal{U}^{erg} . We conclude in the same way as in the proof of Theorem 3.2, applying [11, Corollary 4.8.9] and the analogs of Corollaries 9.3 and 9.4 in [16] for \mathcal{U}^{erg} -valued Ξ -Fleming-Viot processes. \square

Proof of Proposition 3.8 (beginning). In the the first part of the proof, we assume $\Xi \in \mathcal{M}_{\text{dust}}$. Let $n \in \mathbb{N}$, $N \geq n$, $\phi \in \hat{\mathcal{C}}_n$. As in [9], we associate with ϕ not only the marked polynomial Φ but also the marked N -polynomial

$$\Phi_N : \hat{\mathbb{U}}_N \rightarrow \mathbb{R}, \quad \chi \mapsto \nu^{N,\chi}\phi.$$

Let \tilde{p}_N be the transition kernel of the Markov chain $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$. Similarly to the proof of Theorem 3.2, we have

$$\tilde{p}_N \Phi_N(\tilde{\chi}_k^N) = \sum_{\sigma \in \mathcal{S}_n} \mathbb{P}(\gamma_n(\sigma_1^N) = \sigma) \int \nu^{N,\chi}(dr dv) \phi(\sigma(\gamma_n(r, v)) + c_N(0, \underline{\mathbf{1}}_n)).$$

We obtain

$$\begin{aligned} & c_N^{-1}(\tilde{p}_N - I)\Phi_N(\chi) \\ &= \mathbb{P}(\gamma_n(\sigma_1^N) = \emptyset) \int \nu^{N,\chi}(dr dv) c_N^{-1}(\phi(r, v + c_N \underline{\mathbf{1}}_N) - \phi(r, v)) \\ &+ \sum_{\sigma \in \mathcal{S}_n \setminus \{\emptyset\}} c_N^{-1} \mathbb{P}(\gamma_n(\sigma_1^N) = \sigma) \int \nu^{N,\chi}(dr dv) (\phi(\sigma(\gamma_n(r, v)) + c_N(0, \underline{\mathbf{1}}_n)) - \phi(r, v)) \end{aligned}$$

for all $\chi \in \hat{\mathbb{U}}_N$. By condition (3.8) and as $\Xi \in \mathcal{M}_{\text{dust}}$,

$$\lim_{N \rightarrow \infty} c_N^{-1} \mathbb{P}(\gamma_1(\sigma_1^N) = \{\{1\}\}) = \lambda_{1, \{\{1\}\}} < \infty.$$

Here we can use Lemma 6.2 or, more simply, that $b_N = \mathbb{P}(\gamma_1(\sigma_1^N) = \{\{1\}\})$ by exchangeability. Exchangeability and the assumption $\lim_{N \rightarrow \infty} c_N = 0$ now imply

$$\mathbb{P}(\gamma_n(\sigma_1^N) = \emptyset) \geq 1 - n \mathbb{P}(\gamma_1(\sigma_1^N) = \{\{1\}\}) \rightarrow 1 \quad (N \rightarrow \infty).$$

Under our assumption that $\Xi \in \mathcal{M}_{\text{dust}}$, we conclude analogously to the proof of Theorem 3.2. Here we use Lemma 6.2 from the present article, Corollaries 9.3 and 9.4 in [16], and we apply Lemma 4 of [3] to $\hat{\Pi}$. \square

6.2. Proofs of Lemmas 6.1 and 6.2. For $N \in \mathbb{N}$ and $n \in [N]$, we define a probability kernel κ_n^N from Δ^N to \mathcal{P}_n , and a probability kernel K_n^N from Δ^N to \mathcal{S}_n .

For $x \in \Delta^N$, let $\kappa_n^N(x, \cdot)$ be the uniform distribution on those partitions in \mathcal{P}_N whose block sizes are given by (any reordering of) the nonzero elements of the sequence $(Nx(\ell), \ell \in \mathbb{N})$. Then we define $\kappa_n^N(x, \cdot)$ as the restriction $\kappa_n^N(x, \cdot) = \gamma_n(\kappa_n^N(x, \cdot))$.

The distribution of $\kappa_n^N(x, \cdot)$ can also be described by the following urn scheme: Consider an urn that contains $Nx(\ell)$ balls of color ℓ for each $\ell \in \mathbb{N}$. Sample n balls without replacement. Then the random partition of $[n]$ where $i, j \in [n]$ are in the same block if and only if the i -th and j -th ball have the same color has distribution $\kappa_n^N(x, \cdot)$. If we modify this urn scheme such that the balls are sampled with replacement, then we obtain the distribution $\kappa_n^\infty(x, \cdot)$, as a comparison with the definition of $\kappa_n^\infty(x, \cdot)$ in Section 4.1 shows.

For $x \in \Delta^N$, let $(y(1), y(2), \dots)$ be the (possibly empty) finite subsequence of $(Nx(i), i \in \mathbb{N})$ that consists of the elements that are greater or equal to 2. Let $K_N^N(x, \cdot)$ be the uniform distribution on those elements of \mathcal{S}_N that consist of disjoint subsets of $[N]$ whose sizes are given by (an arbitrary reordering of) $(y(1), y(2), \dots)$. If $(Nx(i), i \in \mathbb{N})$ contains no elements that are greater or equal to 2, then $K_N^N(x, \cdot)$ is the Dirac measure in $\emptyset \in \mathcal{S}_N$. We set $K_n^N(x, \cdot) = \gamma_n(K_N^N(x, \cdot))$.

In other words, consider an urn that contains $Nx(\ell)$ balls of color ℓ for each $\ell \in \mathbb{N}$. Recolor each ball whose color occurs only once with a new color 0. Then sample n balls without replacement. Consider the random element σ of \mathcal{S}_n where any $i, j \in [n]$ are in a common block if and only if the i -th and j -th ball have the same color that is not 0. Then σ has distribution $K_n^N(x, \cdot)$. When we sample with replacement, this distribution is $K_n^\infty(x, \cdot)$ instead, as a comparison with the definition of $K_n^\infty(x, \cdot)$ in Section 4.2 shows.

In the following lemma and its proof, we recall and slightly extend some well-known continuity properties from e.g. Proposition 2.9 in Bertoin [1]. We endow \mathbb{N} with the discrete topology and consider the one-point compactification $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Let $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ be the one-point compactification of the space \mathbb{N} , endowed with the discrete topology. We write $\Delta^\infty = \Delta$ and for $\varepsilon > 0$, we define

$$S_\varepsilon = \{(x, N) \in \Delta_\varepsilon \times \bar{\mathbb{N}} : x \in \Delta^N\}.$$

Lemma 6.4. *Let $n \in \mathbb{N}$, $\pi \in \mathcal{P}_n$, $\sigma \in \mathcal{S}_n$, and $\varepsilon > 0$. Then the maps*

$$S_\varepsilon \rightarrow [0, 1], \quad (x, N) \mapsto \kappa_n^N(x, \pi)$$

and

$$S_\varepsilon \rightarrow [0, 1], \quad (x, N) \mapsto K_n^N(x, \sigma)$$

are continuous.

Proof. Let $N \geq n$ and $x \in \Delta^N$. We couple the probability distributions $\kappa_n^\infty(x, \cdot)$ and $\kappa_n^N(x, \cdot)$ by starting with the urn scheme for $\kappa_n^\infty(x, \infty)$ given above and accepting the sample with replacement as a sample without replacement on the event that all sampled balls are different. The probability of this event is bounded from below by $(1 - n/N)^n$, and we obtain the bound

$$|\kappa_n^N(x, \pi) - \kappa_n^\infty(x, \pi)| \leq 1 - (1 - n/N)^n \leq n^2/N.$$

Using the analogous coupling of the probability distributions $K_n^\infty(x, \cdot)$ and $K_n^N(x, \cdot)$, we obtain the bound

$$|K_n^N(x, \sigma) - K_n^\infty(x, \sigma)| \leq 1 - (1 - n/N)^n \leq n^2/N.$$

Furthermore, for $x, y \in \Delta$, we can couple the probability distributions $\kappa_n^\infty(x, \cdot)$ and $\kappa_n^\infty(y, \cdot)$ by using the same uniform random variables in Kingman's correspondence which we recalled in Section 4.1. This yields

$$|\kappa_n^\infty(x, \pi) - \kappa_n^\infty(y, \pi)| \leq n|x - y|_1.$$

Similarly, using the definition of K_n^∞ from Section 4.2, we obtain

$$|K_n^\infty(x, \sigma) - K_n^\infty(y, \sigma)| \leq n|x - y|_1.$$

W.l.o.g., let $((x_k, N_k), k \in \mathbb{N})$ be a sequence in S_ε that converges to some $(x, \infty) \in S_\varepsilon$. Then,

$$\begin{aligned} & |K_n^{N_k}(x_k, \sigma) - K_n^\infty(x, \sigma)| \\ & \leq |K_n^{N_k}(x_k, \sigma) - K_n^\infty(x_k, \sigma)| + |K_n^\infty(x_k, \sigma) - K_n^\infty(x, \sigma)| \\ & \leq n^2/N_k + n|x_k - x|_1, \end{aligned}$$

and the right-hand side converges to zero as k tends to infinity. The argument for $|\kappa_n^{N_k}(x_k, \pi) - \kappa_n^\infty(x, \pi)|$ is the same. \square

Proof of Lemma 6.1. Let $n \in \mathbb{N}$. By construction, we have

$$\mathbb{P}(\gamma_n(\pi_1^N) = \pi) = \int \kappa_n^N(x, \pi) \Xi^N(dx)$$

for all $\pi \in \mathcal{P}_n$ and $N \geq n$. For arbitrarily small $\varepsilon > 0$, assumption (3.7) implies the weak convergence

$$c_N^{-1} \Xi^N(dx) \delta_N(dN') \xrightarrow{w} |x|_2^{-2} \Xi(dx) \delta_\infty(dN') \quad \text{on } \Delta_\varepsilon \times \bar{\mathbb{N}} \quad (N \rightarrow \infty)$$

which also holds on S_ε as none of these measures have mass on the complement of S_ε in $\Delta_\varepsilon \times \bar{\mathbb{N}}$. Using Lemma 6.4, we obtain for each $\pi \in \mathcal{P}_n$

$$\begin{aligned} (6.3) \quad & \lim_{N \rightarrow \infty} c_N^{-1} \int_{\Delta_\varepsilon} \kappa_n^N(x, \pi) \Xi^N(dx) = \lim_{N \rightarrow \infty} \int_{S_\varepsilon} \kappa_n^{N'}(x, \pi) c_N^{-1} \Xi^N(dx) \delta_N(dN') \\ & = \int_{S_\varepsilon} \kappa_n^{N'}(x, \pi) |x|_2^{-2} \Xi(dx) \delta_\infty(dN') = \int_{\Delta_\varepsilon} \kappa_n^\infty(x, \pi) |x|_2^{-2} \Xi(dx). \end{aligned}$$

For every sufficiently small $\varepsilon > 0$, for $N \geq n$, and $x \in \Delta^N \setminus \Delta_\varepsilon$, the urn scheme for the probability distribution $\kappa_n^N(x, \cdot)$ yields that

$$\begin{aligned} (6.4) \quad & \kappa_2^N(x, \{\{1, 2\}\})(1 - n\varepsilon)^n \\ & \leq \kappa_2^N(x, \{\{1, 2\}\}) \frac{N - \varepsilon N}{N - 2} \cdots \frac{N - (n - 2)\varepsilon N}{N - (n - 1)} \\ & \leq \kappa_n^N(x, \{\{1, 2\}, \{3\}, \dots, \{n\}\}) \\ & \leq \kappa_2^N(x, \{\{1, 2\}\}). \end{aligned}$$

Indeed, we can draw a partition of $\{1, \dots, n\}$ according to the distribution $\kappa_n^N(x, \cdot)$ by drawing n times without replacement from N individuals that are subdivided into families of sizes $(Nx(\ell), \ell \in \mathbb{N})$, and letting i and j be in the same block if and only if the i -th and j -th drawn individual are in the same family. The second inequality follows as the family sizes are at most εN by our assumption on x .

For $x \in \Delta \setminus \Delta_\varepsilon$ and every partition $\pi \in \mathcal{P}_n$ that contains one doubleton and apart from that only singletons, we have

$$(6.5) \quad \kappa_n^N(x, \pi) = \kappa_n^N(x, \{\{1, 2\}, \{3\}, \dots, \{n\}\})$$

by exchangeability. Moreover,

$$(6.6) \quad c_N = \int \kappa_2^N(x, \{\{1, 2\}\}) \Xi^N(dx)$$

by exchangeability.

Consequently, for arbitrarily small $\varepsilon > 0$ as in condition (3.7), and any partition $\pi \in \mathcal{P}_n$ that contains one doubleton and apart from that only singletons,

$$\begin{aligned} & \limsup_{N \rightarrow \infty} c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} \kappa_n^N(x, \pi) \Xi^N(dx) \leq \limsup_{N \rightarrow \infty} c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} \kappa_2^N(x, \{\{1, 2\}\}) \Xi^N(dx) \\ & = 1 - \liminf_{N \rightarrow \infty} c_N^{-1} \int_{\Delta_\varepsilon} \kappa_2^N(x, \{\{1, 2\}\}) \Xi^N(dx) = 1 - \Xi(\Delta_\varepsilon), \end{aligned}$$

where we use (6.4) and (6.5) for the first, equation (6.6) for the second, and the convergence (6.3) and $\kappa_2^\infty(x, \{\{1, 2\}\}) = |x|_2^2$ for the third step. Analogously, we obtain for the same partitions π that

$$\liminf_{N \rightarrow \infty} c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} \kappa_n^N(x, \pi) \Xi^N(dx) \geq (1 - n\varepsilon)^n (1 - \Xi(\Delta_\varepsilon)).$$

For every partition $\pi \in \mathcal{P}_n$ that contains a block of size greater than 2, $N \geq n$, and $x \in \Delta^N \setminus \Delta_\varepsilon$ we have

$$\kappa_n^N(x, \pi) \leq \kappa_3^N(x, \{\{1, 2, 3\}\}) \leq \kappa_2^N(x, \{\{1, 2\}\}) \frac{N\varepsilon - 2}{N - 2}$$

which can be seen from the urn scheme for $\kappa_n^N(x, \cdot)$ and the exchangeability therein. We obtain for these partitions π that

$$\limsup_{N \rightarrow \infty} c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} \kappa_n^N(x, \pi) \Xi^N(dx) \leq \varepsilon(1 - \Xi(\Delta_\varepsilon)).$$

Similarly, for every partition $\pi \in \mathcal{P}_n$ that contains more than one non-singleton block, $N \geq n$, and $x \in \Delta^N \setminus \Delta_\varepsilon$, we have

$$\kappa_n^N(x, \pi) \leq \kappa_4^N(x, \{\{1, 2\}, \{3, 4\}\}) \leq \kappa_2^N(x, \{\{1, 2\}\}) \frac{N\varepsilon - 1}{N - 3},$$

hence

$$\limsup_{N \rightarrow \infty} c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} \kappa_n^N(x, \pi) \Xi^N(dx) \leq \varepsilon(1 - \Xi(\Delta_\varepsilon)).$$

For $\pi \in \mathcal{P}_n \setminus \{\mathbf{0}_n\}$, we write

$$c_N^{-1} \mathbb{P}(\gamma_n(\pi_1^N) = \pi) = c_N^{-1} \int_{\Delta_\varepsilon} \kappa_n^N(x, \pi) \Xi^N(dx) + c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} \kappa_n^N(x, \pi) \Xi^N(dx).$$

Now we let first $N \rightarrow \infty$. For the first integral, we use the convergence (6.3). For the second integral, we use the bounds for the limes superior and if necessary also for the limes inferior. Then we let ε tend to zero along a sequence such that for each element of this sequence, the weak convergence in condition (3.7) holds. Thus we obtain convergence to the rates λ_π as asserted in Lemma 6.1. \square

Proof of Lemma 6.2. The proof is analogous to Lemma 6.1. By construction, we have

$$\mathbb{P}(\gamma_n(\sigma_1^N) = \sigma) = \int K_n^N(x, \sigma) \Xi^N(dx)$$

for all $\sigma \in \mathcal{S}_n$ and $N \geq n$. Analogously to the convergence (6.3), we obtain from Lemma 6.4 and assumption (3.7) that

$$(6.7) \quad \lim_{N \rightarrow \infty} c_N^{-1} \int_{\Delta_\varepsilon} K_n^N(x, \sigma) \Xi^N(dx) = \int_{\Delta_\varepsilon} K_n^\infty(x, \sigma) |x|_2^{-2} \Xi(dx)$$

for all $\sigma \in \mathcal{S}_n$ and arbitrarily small $\varepsilon > 0$.

For every $\sigma \in \mathcal{S}_n \setminus \{\emptyset\}$, $N \geq n$, and $x \in \Delta^N$, it holds

$$K_n^N(x, \sigma) \leq K_1^N(x, \{\{1\}\}),$$

by exchangeability. Furthermore,

$$b_N = \int K_1^N(x, \{\{1\}\}) \Xi^N(dx).$$

Now we obtain for these σ and arbitrarily small $\varepsilon > 0$

$$(6.8) \quad \begin{aligned} & \limsup_{N \rightarrow \infty} c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} K_n^N(x, \sigma) \Xi^N(dx) \leq \limsup_{N \rightarrow \infty} c_N^{-1} \int_{\Delta \setminus \Delta_\varepsilon} K_1^N(x, \{\{1\}\}) \Xi^N(dx) \\ & = \lim_{N \rightarrow \infty} b_N / c_N - \lim_{N \rightarrow \infty} c_N^{-1} \int_{\Delta_\varepsilon} K_1^N(x, \{\{1\}\}) \Xi^N(dx) = \int_{\Delta \setminus \Delta_\varepsilon} |x|_1 |x|_2^{-2} \Xi(dx), \end{aligned}$$

where we use assumption (3.8), the convergence (6.7), and $K_1^\infty(x, \{\{1\}\}) = |x|_1$ in the last equality. We conclude by combining the convergence (6.7) and the bound (6.8). We let ε tend to zero and use dominated convergence as the integrands on the right-hand sides of (6.7) and (6.8) are bounded by $|x|_1|x|_2^{-2}$, and as

$$\int_{\Delta} |x|_1 |x|_2^{-2} \Xi(dx) < \infty.$$

This yields the assertion. \square

7. CONVERGENCE OF MARKED METRIC MEASURE SPACES IN THE DUST-FREE CASE

In this section, we prove Theorem 3.5 and Proposition 3.8 for $\Xi \in \mathcal{M}_{\text{nd}}$. We use the isometric embedding $\beta_0 : \mathbb{U}_N \rightarrow \hat{\mathbb{U}}_N$, $\llbracket X, \rho, \mu \rrbracket \mapsto \llbracket X, \rho, \mu \otimes \delta_0 \rrbracket$ which maps a metric measure space to the associated marked metric measure space that supports only the zero mark. We compare the process $(\beta_0(\chi_{[c_N^{-1}t]}^N), t \in \mathbb{R}_+)$ to the processes $(\hat{\chi}_{[c_N^{-1}t]}^N, t \in \mathbb{R}_+)$ and $(\tilde{\chi}_{[c_N^{-1}t]}^N, t \in \mathbb{R}_+)$.

Recall from Section 3.4 the map $\beta : \mathbb{U}_N \rightarrow \hat{\mathbb{U}}_N$ which decomposes a tree at the external branches. Also recall from Section 2.1 the functions α and Υ . We denote also by α the function from $\hat{\mathbb{U}}_N$ to \mathbb{U}_N that maps χ to $\psi_N(\alpha(r, v))$, where (r, v) is any element of $\hat{\mathbb{U}}_N$ with $\hat{\psi}_N(r, v) = \chi$. The function $\alpha : \hat{\mathbb{U}}_N \rightarrow \mathbb{U}_N$ retrieves a metric measure space from a marked metric measure space by adding the marks to the metric distances. For $\ell \geq 2$, we define the map

$$\Upsilon_1^\ell : \mathfrak{U} \rightarrow \mathbb{R}_+, \quad \Upsilon_1^\ell = \gamma_1 \circ \Upsilon \circ \gamma_\ell.$$

In the subtree spanned by the first ℓ leaves of the tree associated with some $\rho \in \mathfrak{U}$, the length of the external branch that ends in the first leaf is given by $\Upsilon_1^\ell(\rho)$. We also define the restriction $\varpi : \hat{\mathfrak{U}} \rightarrow \mathbb{R}_+$, $(r, v) \mapsto v(1)$.

We recall that the Prohorov metric d_P on the space of probability measures on the Borel sigma algebra of a metric space (S, d) is given by

$$(7.1) \quad d_P(\mu, \mu') = \inf\{\varepsilon > 0 : \mu'(F) \leq \mu(F^\varepsilon) + \varepsilon \text{ for all closed } F \subset S\},$$

where $F^\varepsilon = \{x \in S : d(x, F) < \varepsilon\}$. If (S, d) is separable, then we also have the coupling characterization

$$(7.2) \quad d_P(\mu, \mu') = \inf_\nu \inf\{\varepsilon > 0 : \nu\{(x, y) \in S^2 : d(x, y) > \varepsilon\} < \varepsilon\}$$

where the first infimum is over all couplings ν of μ and μ' , see e. g. [11, Theorem 3.1.2].

Lemma 7.1. *Let $\chi \in \hat{\mathbb{U}}_N$. Then,*

$$d_{\text{mGP}}(\chi, \beta_0 \circ \alpha(\chi)) \leq d_P(\varpi(\nu^\chi), \delta_0).$$

Corollary 7.2. *For every $\chi \in \mathbb{U}_N$,*

$$d_{\text{mGP}}(\beta(\chi), \beta_0(\chi)) \leq d_P(\varpi(\nu^{\beta(\chi)}), \delta_0).$$

Proof. This is immediate from Lemma 7.1 as $\alpha \circ \beta$ is the identity on \mathbb{U}_N . \square

We prove Lemma 7.1 using a characterization of the marked Gromov-Prohorov metric d_{mGP} that we will apply also in Section 8. The distortion $\text{dis } \mathfrak{R}$ of a relation $\mathfrak{R} \subset X \times X'$ between two metric spaces (X, r) and (X', r') is defined by

$$\text{dis } \mathfrak{R} = \sup\{|r(x, y) - r'(x', y')| : (x, x'), (y, y') \in \mathfrak{R}\}.$$

Proposition 7.3. *Let (X, r, m) and (X', r', m') be marked metric measure spaces. Then $d_{\text{mGP}}(\llbracket X, r, m \rrbracket, \llbracket X', r', m' \rrbracket)$ is the infimum of all $c > 0$ such that there exist a relation $\mathfrak{R} \subset X \times X'$ and a coupling ν of m and m' with $\frac{1}{2}\text{dis } \mathfrak{R} \leq c$ and $\nu(\tilde{\mathfrak{R}}) \geq 1 - c$, where $\hat{\mathfrak{R}} \subset (X \times \mathbb{R}_+) \times (X' \times \mathbb{R}_+)$ is defined by*

$$\hat{\mathfrak{R}} = \{((x, v), (x', v')) : (x, x') \in \mathfrak{R}, |v - v'| \leq c\}.$$

Proof. This can be seen as an adaptation of Proposition 6 in [20]. Here we sketch the proof of the upper bound for $d_{\text{mGP}}(\llbracket X, r, m \rrbracket, \llbracket X', r', m' \rrbracket)$. Let $c > 0$ and assume \mathfrak{R} , $\hat{\mathfrak{R}}$, and ν with $\frac{1}{2}\text{dis } \mathfrak{R} \leq c$ and $\nu(\hat{\mathfrak{R}}) \geq 1 - c$ are given as in the proposition. A metric d^Z on the disjoint union $Z = X \sqcup X'$ can be defined by $d^Z(x, y) = r(x, y)$ for $(x, y) \in X$, $d^Z(x, y) = r'(x, y)$ for $(x, y) \in X'$, and

$$d^Z(x, x') = \inf\{r(x, y) + c + r'(y', x') : (y, y') \in \mathfrak{R}\},$$

for $x \in X, x' \in X'$, as in Remark 5.5 of [12]. We endow $Z \times \mathbb{R}_+$ with the product metric $d^{Z \times \mathbb{R}_+}((z, v), (z', v')) = d^Z(z, z') \vee |v - v'|$. Let $\varphi : X \rightarrow Z$ and $\varphi' : X' \rightarrow Z$ be the canonical embeddings. Moreover, let $\hat{\varphi}(x, v) = (\varphi(x), v)$ and $\hat{\varphi}'(x', v) = (\varphi'(x'), v)$ for $x \in X, x' \in X'$, and $v \in \mathbb{R}_+$. Then the coupling ν induces a coupling $\hat{\nu}$ of $\hat{\varphi}(m)$ and $\hat{\varphi}'(m')$ on $Z \times \mathbb{R}_+$ with

$$\hat{\nu}\{((z, v), (z', v')) : d^Z(z, z') \vee |v - v'| \leq c\} \geq 1 - c.$$

The coupling characterization (7.2) of the Prohorov metric implies $d_P(\hat{\varphi}(m), \hat{\varphi}'(m')) \leq c$. The definition of the marked Gromov-Prohorov metric, see [9, Definition 3.1], implies

$$d_{\text{mGP}}(\llbracket X, r, m \rrbracket, \llbracket X', r', m' \rrbracket) \leq c.$$

□

Proof of Lemma 7.1. Let (r, v) be any element of $\hat{\mathfrak{U}}_N$ with $\hat{\psi}_N(r, v) = \chi$, and let $\rho = \alpha(r, v)$. Then,

$$\llbracket [N], r, N^{-1} \sum_{i=1}^N \delta_{(i, v(i))} \rrbracket = \chi$$

and

$$\llbracket [N], \rho, N^{-1} \sum_{i=1}^N \delta_{(i, 0)} \rrbracket = \beta_0 \circ \alpha(\chi).$$

Let $c > d_P(\varpi(\nu^\chi), \delta_0)$. We conclude by Proposition 7.3 which also holds for marked semi-metric measure spaces. To this aim, we define the relation

$$\mathfrak{R} = \{(i, i) \in [N] \times [N] : v(i) \leq c\}$$

between the semi-metric spaces $([N], r)$ and $([N], \rho)$. As $|r(i, j) - \rho(i, j)| \leq v(i) + v(j)$ by definition of the map $\alpha : \hat{\mathfrak{U}}_N \rightarrow \mathfrak{U}_N$, we can bound the distortion by

$$\text{dis } \mathfrak{R} = \max\{|r(i, j) - \rho(i, j)| : i, j \in [N], v(i), v(j) \leq c\} \leq 2c.$$

We set

$$\hat{\mathfrak{R}} = \{((i, v(i)), (i, 0)) : i \in [N], v(i) \leq c\} \subset ([N] \times \mathbb{R}_+) \times ([N] \times \mathbb{R}_+).$$

A coupling ν of the probability measures $N^{-1} \sum_{i=1}^N \delta_{(i, v(i))}$ and $N^{-1} \sum_{i=1}^N \delta_{(i, 0)}$ is given by

$$\nu = N^{-1} \sum_{i=1}^N \delta_{((i, v(i)), (i, 0))}.$$

Finally,

$$\nu(\hat{\mathfrak{R}}) = N^{-1} \sum_{i=1}^N \mathbf{1}\{v(i) \leq c\} = \varpi(\nu^\chi)[0, c] \geq \delta_0\{0\} - c = 1 - c,$$

where the inequality follows from the choice of c and the usual definition (7.1) of the Prohorov metric. □

Lemma 7.4. *Let $\Xi \in \mathcal{M}_{\text{nd}}$, and let $(\chi_t, t \in \mathbb{R}_+)$ be a \mathbb{U} -valued Ξ -Fleming-Viot process with càdlàg paths. Then,*

$$\lim_{\ell \rightarrow \infty} \sup_{t \in [0, T]} d_P(\Upsilon_1^\ell(\nu^{\chi_t}), \delta_0) = 0 \quad \text{a. s.}$$

for all $T \in \mathbb{R}_+$.

Proof. Let $\varepsilon > 0$. For $\ell \geq 2$, we define the random time

$$\vartheta_{\varepsilon, \ell} = \inf\{t \in \mathbb{R}_+ : d_P(\Upsilon_1^\ell(\nu^{\chi_t}), \delta_0) > \varepsilon\}.$$

For all $\rho \in \mathfrak{U}$, the map $\ell \mapsto \Upsilon_1^\ell(\rho)$ is non-increasing. We set

$$\vartheta_\varepsilon = \sup_{\ell \in \mathbb{N}} \vartheta_{\varepsilon, \ell} = \lim_{\ell \rightarrow \infty} \vartheta_{\varepsilon, \ell}.$$

Let $t \in (0, \infty)$ be arbitrary. On an event of probability 1, let (X, ρ, μ) be a representative of $\chi_{\vartheta_\varepsilon \wedge t}$, and let $(x_i, i \in \mathbb{N})$ be a μ -iid sequence in X . Then,

$$\inf_{j \in \mathbb{N} \setminus \{1\}} \rho(x_1, x_j) = 0 \quad \text{a. s.}$$

as an iid sequence with respect to a probability measure on the Borel sigma algebra on a separable metric space has a.s. no isolated elements. Consequently,

$$\lim_{\ell \rightarrow \infty} \Upsilon_1^\ell((\rho(x_i, x_j))_{i, j \in \mathbb{N}}) = 0 \quad \text{a. s.}$$

As the random matrix $(\rho(x_i, x_j))_{i, j \in \mathbb{N}}$ has conditional distribution $\nu^{\chi_{\vartheta_\varepsilon \wedge t}}$ given $\chi_{\vartheta_\varepsilon \wedge t}$, it follows

$$\lim_{\ell \rightarrow \infty} d_P(\Upsilon_1^\ell(\nu^{\chi_{\vartheta_\varepsilon \wedge t}}), \delta_0) = 0 \quad \text{a. s.}$$

Analogously, it can be shown that

$$\lim_{\ell \rightarrow \infty} d_P(\Upsilon_1^\ell(\nu^{\chi_{(\vartheta_\varepsilon \wedge t)^-}}), \delta_0) = 0 \quad \text{a. s.}$$

As $(\chi_s, s \in \mathbb{R}_+)$ has càdlàg paths and as the maps $\chi \mapsto \nu^\chi$ and Υ_1^ℓ are continuous, it follows that on an event of probability 1, there exists $\ell \in \mathbb{N}$ such that

$$d_P(\Upsilon_1^\ell(\nu^{\chi_s}), \delta_0) < \varepsilon/2$$

for all s in a neighborhood of $\vartheta_\varepsilon \wedge t$. By monotonicity, it also holds

$$d_P(\Upsilon_1^{\ell'}(\nu^{\chi_s}), \delta_0) < \varepsilon/2$$

for all $\ell' \geq \ell$ and s in the same neighborhood, on the same event of probability 1. This implies $\vartheta_{\ell, \varepsilon} > t$ for ℓ sufficiently large a.s., hence $\{\vartheta_\varepsilon < t\}$ is a null event. As t was arbitrary, it follows $\vartheta_\varepsilon = \infty$ a.s. which yields the assertion. \square

Proof of Theorem 3.5 (beginning). First we assume $\Xi \in \mathcal{M}_{\text{nd}}$. Let $T \in \mathbb{R}_+$. By the assumption that $\hat{\chi}_0$ supports only the zero mark, there exists $\chi_0 \in \mathbb{U}$ such that $\hat{\chi}_0^N$ converges to $\beta_0(\chi_0)$ in the marked Gromov-weak topology as N tends to infinity. Hence, $\nu^{\hat{\chi}_0^N}$ converges weakly to $\nu^{\chi_0} \otimes \delta_0$. Recall the chain $(\chi_k^N, k \in \mathbb{N}_0)$ from Section 3.2. As $\nu^{\chi_0^N} = \alpha(\nu^{\hat{\chi}_0^N})$ converges weakly to ν^{χ_0} , also χ_0^N converges to χ_0 in the Gromov-weak topology. Hence, Theorem 3.2 is applicable and the processes $(\chi_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+)$ converge in distribution to a \mathbb{U} -valued Ξ -Fleming-Viot process $(\chi_t, t \in \mathbb{R}_+)$ with initial state χ_0 in the space of càdlàg paths in $(\mathbb{U}, d_{\text{GP}})$, endowed with the Skorohod metric. For every $\ell \geq 2$, by continuity of the maps $\chi \mapsto \nu^\chi$ and Υ_1^ℓ , also $(\Upsilon_1^\ell(\nu^{\chi_{\lfloor c_N^{-1}t \rfloor}^N}), t \in \mathbb{R}_+)$ converges in distribution to $(\Upsilon_1^\ell(\nu^{\chi_t}), t \in \mathbb{R}_+)$ in the space of càdlàg paths in $(\mathcal{M}_1(\mathbb{R}_+), d_P)$, endowed with the Skorohod metric, where $\mathcal{M}_1(\mathbb{R}_+)$ denotes the space of probability measures on \mathbb{R}_+ .

For every $(r, v) \in \hat{\mathfrak{U}}$ and $\rho = \alpha(r, v)$, it holds $\rho(1, j) \geq v(1)$ for all $j \geq 2$ by definition of the map $\alpha : \hat{\mathfrak{U}}_N \rightarrow \mathfrak{U}_N$, hence $v(1) \leq 2\Upsilon_1^\ell(\alpha(r, v))$ for all $\ell \geq 2$. This implies

$$\begin{aligned} (7.3) \quad & \sup_{t \in [0, T]} d_P(\varpi(\nu^{\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}), \delta_0) \\ & \leq \sup_{t \in [0, T]} d_P(2\Upsilon_1^\ell(\alpha(\nu^{\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N))), \delta_0) \\ & \leq 2 \sup_{t \in [0, T]} d_P(\Upsilon_1^\ell(\nu^{\chi_{\lfloor c_N^{-1}t \rfloor}^N}), \delta_0). \end{aligned}$$

The expression in the last line converges in distribution to

$$(7.4) \quad 2 \sup_{t \in [0, T]} d_P(\Upsilon_1^\ell(\nu^{\chi_t}), \delta_0)$$

as N tends to infinity. This follows from the discussion in the beginning of this proof, as the maps Υ_1^ℓ and $d_P(\cdot, \delta_0)$ are continuous, and as the process $(\Upsilon_1^\ell(\nu^{\chi_t}), t \in \mathbb{R}_+)$ has a.s. no discontinuity at the fixed time T . Finally, we let ℓ tend to infinity. By Lemma 7.4, expression (7.4) then converges to zero a.s. Consequently, also the left-hand side of (7.3) converges to zero in probability as N tends to infinity.

As $\hat{\chi}_k^N = \beta(\chi_k^N)$ for all $k \in \mathbb{N}_0$, Corollary 7.2 implies that

$$\sup_{t \in [0, T]} d_{\text{mGP}}(\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, \beta_0(\chi_{\lfloor c_N^{-1}t \rfloor}^N))$$

converges to zero in probability. The processes $(\beta_0(\chi_{\lfloor c_N^{-1}t \rfloor}^N), t \in \mathbb{R}_+)$ converge in distribution to $(\beta_0(\chi_t), t \in \mathbb{R}_+)$ by another application of Theorem 3.2. The assertion for $\Xi \in \mathcal{M}_{\text{nd}}$ now follows from Slutsky's theorem. \square

To prove Proposition 3.8 in case $\Xi \in \mathcal{M}_{\text{nd}}$, we will use the following coupling of $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$ and $(\chi_k^N, k \in \mathbb{N}_0)$.

Remark 7.5. If $\chi_0^N = \alpha(\tilde{\chi}_0^N)$, then we can define (r_0^N, v_0^N) and ρ_0^N in Section 3 such that

$$\rho_0^N = \alpha(r_0^N, v_0^N).$$

We can then also assume that the processes $((r_k^N, v_k^N), k \in \mathbb{N}_0)$, $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$, and $(\rho_k^N, k \in \mathbb{N}_0)$ are defined as in Section 3.5. Then we obtain $\rho_k^N = \alpha(r_k^N, v_k^N)$ for all $k \in \mathbb{N}$. When we define $\chi_k^N = \psi_N(\rho_k^N)$ and $\hat{\chi}_k^N = \hat{\psi}_N \circ \beta(\rho_k^N)$ as in Section 3, then we also have $\chi_k^N = \alpha(\tilde{\chi}_k^N)$ and $\hat{\chi}_k^N = \beta \circ \alpha(\tilde{\chi}_k^N)$ for all $k \in \mathbb{N}_0$.

Proof of Proposition 3.8 (end). In case $\Xi \in \mathcal{M}_{\text{nd}}$, the proof is almost identical with the proof of Theorem 3.5. We replace $\hat{\chi}_\cdot^N$ with $\tilde{\chi}_\cdot^N$, and we set $\chi_0^N = \alpha(\tilde{\chi}_0^N)$. Then we use the coupling from Remark 7.5 and apply Theorem 3.2. \square

8. CONVERGENCE OF MARKED METRIC MEASURE SPACES IN THE CASE WITH DUST

In this section, we complete the proof of Theorem 3.5: In the case with dust, we compare the finite dimensional distributions of the processes $(\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+)$ and $(\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+)$, and we show relative compactness of the sequence of processes $((\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+), N \geq 2)$.

A metric \tilde{d} on $\hat{\mathcal{U}}_N$ is defined by

$$\tilde{d}((r, v), (r', v')) = \max_{i, j \in [N]} |r(i, j) - r'(i, j)| \vee \max_{i \in [N]} |v(i) - v'(i)|.$$

W.l.o.g., we endow $\hat{\mathcal{U}}_N$ with the metric

$$d((r, v), (r', v')) = \max_{n \in [N]} (\tilde{d}(\gamma_n(r, v), \gamma_n(r', v')) \wedge (2^{-n}))$$

and $\hat{\mathcal{U}}$ with the metric

$$d((r, v), (r', v')) = \sup_{n \in \mathbb{N}} (\tilde{d}(\gamma_n(r, v), \gamma_n(r', v')) \wedge (2^{-n})).$$

These metrics induce the topologies on $\hat{\mathcal{U}}_N$ and \mathcal{U} from Section 2.1. We have

$$(8.1) \quad d(\gamma_n(r, v), \gamma_n(r', v')) \leq d((r, v), (r', v')) \leq d(\gamma_n(r, v), \gamma_n(r', v')) + 2^{-n}$$

for all $(r, v), (r', v') \in \hat{\mathcal{U}}$.

We use definitions in particular from Section 2.1 and we need the following lemmas.

Lemma 8.1. Let $n \in \mathbb{N}$ and $(r, v), (r', v') \in \hat{\mathfrak{U}}_n$ with $\alpha(r, v) = \alpha(r', v')$. Then,

$$d((r, v), (r', v')) \leq 2 \max_{i \in [n]} |v(i) - v'(i)|.$$

Proof. Let $\rho = \alpha(r, v) = \alpha(r', v')$. By definition of α ,

$$|r(i, j) - r'(i, j)| = |\rho(i, j) - v(i) - v(j) - \rho(i, j) + v'(i) + v'(j)| \leq |v(i) - v'(i)| + |v(j) - v'(j)|$$

for all distinct $i, j \in [n]$. It follows

$$d((r, v), (r', v')) \leq \tilde{d}((r, v), (r', v')) \leq 2 \max_{i \in [n]} |v(i) - v'(i)|.$$

□

In the next lemma, we consider the chain $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$ from Section 3.5. We show that if the initial state corresponds to the decomposition at the external branches, then in all generations, the mark of a sampled individual is not larger than the length of the corresponding external branch.

Lemma 8.2. Let the $\hat{\mathfrak{U}}_N$ -valued chain $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$ be defined from the $\hat{\mathfrak{U}}_N$ -valued chain $((r_k^N, v_k^N), k \in \mathbb{N}_0)$ as in Section 3.5, and assume that $(r_0^N, v_0^N) = \beta \circ \alpha(r_0^N, v_0^N)$. Then,

$$v \leq \Upsilon \circ \alpha(r, v) \quad (\text{component-wise})$$

for $\nu^{N, \tilde{\chi}_k^N}$ -a. a. $(r, v) \in \hat{\mathfrak{U}}_N$ and all $k \in \mathbb{N}_0$ a. s.

Proof. Let $(\rho_k^N, k \in \mathbb{N}_0) = (\alpha(r_k^N, v_k^N), k \in \mathbb{N}_0)$ as in Section 3.5. By definition of Υ ,

$$(8.2) \quad \Upsilon(\rho_k^N)(i) = \frac{1}{2} \min_{j \in [N] \setminus \{i\}} \rho_k^N(i, j)$$

for all $i \in [N]$ and $k \in \mathbb{N}_0$. We show $v_k^N(i) \leq \Upsilon(\rho_k^N)(i)$. The assertion follows as for $\nu^{N, \tilde{\chi}_k^N}$ -a. a. $(r, v) \in \hat{\mathfrak{U}}_N$, there exists a.s. a bijection p on $[N]$ with $v(i) = v_k^N(p(i))$ for all $i \in [N]$ by definition of the N -marked distance matrix distribution.

Let $i \in [N]$, $k \in \mathbb{N}_0$, and $\bar{v}(i) = \Upsilon(\rho_k^N)(i)$. To show $v_k^N(i) \leq \bar{v}(i)$, we consider the cases $v_k^N(i) \leq c_N k$ and $v_k^N(i) \geq c_N k$ separately.

In the case $v_k^N(i) \leq c_N k$, there are, by definition of v_k^N , no reproduction events from generations $k - c_N^{-1} v_k^N(i) + 1$ to k due to which the ancestral lineage of the individual i of generation k can merge with the ancestral lineage of a different individual. Hence, by definition of ρ_k^N in Section 3.1,

$$\min_{j \in [N] \setminus \{i\}} \rho_k^N(i, j) \geq 2v_k^N(i).$$

Equation (8.2) yields $v_k^N(i) \leq \bar{v}(i)$ in case $v_k^N(i) \leq c_N k$.

The statements in the remainder of this proof hold in the case $v_k^N(i) \geq c_N k$. There are no reproduction events in which the ancestral lineage of the individual i of generation k can merge with the ancestral lineage of a different individual. Hence, $A_0(k, i) \neq A_0(k, j)$ for all $j \in [N] \setminus \{i\}$. By definition of ρ_k^N ,

$$(8.3) \quad \rho_k^N(i, j) = 2c_N k + \rho_0^N(A_0(k, i), A_0(k, j))$$

for all $j \in [N] \setminus \{i\}$. By definition of v_k^N ,

$$v_k^N(i) = c_N k + v_0^N(A_0(k, i)).$$

Also, $v_0^N = \Upsilon(\rho_0^N)$ by our assumption, hence

$$(8.4) \quad v_0^N(A_0(k, i)) \leq \frac{1}{2} \rho_0^N(A_0(k, i), \ell)$$

for all $\ell \in [N] \setminus \{A_0(k, i)\}$. Using equations (8.2), (8.3), and (8.4), we obtain

$$\bar{v}(i) = \frac{1}{2} \min_{j \in [N] \setminus \{p(i)\}} \rho_k^N(i, j) \geq c_N k + v_0^N(A_0(k, i)) = v_k^N(i).$$

□

For $\ell \geq n \geq 2$, we introduce the map

$$\Upsilon_n^\ell : \mathfrak{U} \cup \bigcup_{N \geq \ell} \mathfrak{U}_N \rightarrow \mathbb{R}_+^n, \quad \Upsilon_n^\ell = \gamma_n \circ \Upsilon \circ \gamma_\ell.$$

The vector $\Upsilon_n^\ell(\rho)$ gives the lengths of the first n external branches in the subtree spanned by the first ℓ leaves of the tree associated with some $\rho \in \mathfrak{U}_N$ or $\rho \in \mathfrak{U}$. We also define the restriction $\varpi_{\mathbb{R}_+^n} : \hat{\mathfrak{U}} \rightarrow \mathbb{R}_+^n$, $(r, v) \mapsto v$. We endow \mathbb{R}_+^n with the maximum norm and the induced metric. Let $\gamma_n : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$ be the restriction $v \mapsto (v(i))_{i \in [n]}$. Recall also the maps α and β from Sections 2.1 and 7.

Proof of Theorem 3.5 (continuation). We set $\tilde{\chi}_0^N = \hat{\chi}_0^N$. Then, $\chi_0^N = \alpha(\tilde{\chi}_0^N)$ and we can assume that the processes $(\tilde{\chi}_k^N, k \in \mathbb{N}_0)$, $(\hat{\chi}_k^N, k \in \mathbb{N}_0)$, and $((r_k^N, v_k^N), k \in \mathbb{N}_0)$ are defined as in Remark 7.5. In particular, we have $\hat{\chi}_k^N = \beta \circ \alpha(\tilde{\chi}_k^N)$ for all $k \in \mathbb{N}_0$ a.s.

First we consider finite-dimensional distributions. Let $t \in \mathbb{R}_+$. For $N \geq \ell \geq n \geq 2$,

$$\begin{aligned} & d_P(\nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}, \nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}) \\ & \leq d_P(\gamma_n(\nu^{\beta \circ \alpha(\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N)}), \gamma_n(\nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) + 2^{-n+1} \\ & \leq d_P(\gamma_n(\nu^{N, \beta \circ \alpha(\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N)}), \gamma_n(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) + 2^{-n+1} + 2n^2/N \\ & = d_P(\gamma_n \circ \beta \circ \alpha(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}), \gamma_n(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) + 2^{-n+1} + 2n^2/N \\ & \leq 2d_P(\gamma_n \circ \Upsilon \circ \alpha(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}), \gamma_n \circ \varpi_{\mathbb{R}_+^n}(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) + 2^{-n+1} + 2n^2/N \quad \text{a.s.} \end{aligned}$$

For the first inequality, we use relation (8.1) and either (7.1) or (7.2). For the last inequality, we use Lemma 8.1, the definitions of Υ and β , and (7.1) or (7.2). For the second inequality, we use the bounds

$$d_P(\gamma_n(\nu^{\beta \circ \alpha(\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N)}), \gamma_n(\nu^{N, \beta \circ \alpha(\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N)})) \leq n^2/N$$

and

$$d_P(\gamma_n(\nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}), \gamma_n(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) \leq n^2/N$$

which can be seen from the coupling characterization (7.2) of the Prohorov metric. Here we couple sampling with and without replacement as in the proofs of e.g. (6.2) or of Lemma 6.4.

By definition of Υ_n^ℓ and Υ ,

$$\gamma_n \circ \Upsilon(\rho) \leq \Upsilon_n^\ell(\rho)$$

for all $\rho \in \mathfrak{U}_N$. Using Lemma 8.2, we obtain

$$|\gamma_n \circ \Upsilon \circ \alpha(r, v) - \gamma_n(v)| \leq |\Upsilon_n^\ell \circ \alpha(r, v) - \gamma_n(v)|$$

for $\nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}$ -a. a. $(r, v) \in \hat{\mathfrak{U}}_N$ a.s. Again using the definition of the Prohorov metric, we obtain the first inequality in the following display.

$$\begin{aligned} & d_P(\gamma_n \circ \Upsilon \circ \alpha(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}), \gamma_n \circ \varpi_{\mathbb{R}_+^n}(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) \\ & \leq d_P(\Upsilon_n^\ell \circ \alpha(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}), \gamma_n \circ \varpi_{\mathbb{R}_+^n}(\nu^{N, \tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) \\ (8.5) \quad & \leq d_P(\Upsilon_n^\ell \circ \alpha(\nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}), \gamma_n \circ \varpi_{\mathbb{R}_+^n}(\nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})) + \ell^2/N + n^2/N \quad \text{a.s.} \end{aligned}$$

For the second inequality, we again couple sampling with and without replacement, and we use the triangle inequality twice. By continuity of the maps $\chi \mapsto \nu^\chi$, α , Υ_n^ℓ , $\varpi_{\mathbb{R}_+^n}$, γ_n , and $d_P(\cdot, \cdot)$, and by Proposition 3.8, the right-hand side of (8.5) converges in distribution to

$$(8.6) \quad d_P(\Upsilon_n^\ell \circ \alpha(\nu^{\hat{\chi}^t}), \gamma_n \circ \varpi_{\mathbb{R}_+^n}(\nu^{\hat{\chi}^t}))$$

as $N \rightarrow \infty$, where $(\hat{\chi}_s, s \in \mathbb{R}_+)$ is a $\hat{\mathbb{U}}$ -valued Ξ -Fleming-Viot process with initial state $\hat{\chi}_0$. Here we also use that $(\hat{\chi}_s, s \in \mathbb{R}_+)$ has a.s. no discontinuity at the fixed time t .

For every $\rho \in \mathfrak{U}$,

$$\lim_{\ell \rightarrow \infty} \Upsilon_n^\ell(\rho) = \gamma_n \circ \Upsilon(\rho)$$

by definition of Υ_n^ℓ , and by the definition of $\Upsilon : \mathfrak{U} \rightarrow \mathbb{R}_+^\mathbb{N}$ in Section 2.1. Hence, for every $\chi \in \hat{\mathbb{U}}$,

$$\lim_{\ell \rightarrow \infty} d_P(\Upsilon_n^\ell \circ \alpha(\nu^\chi), \gamma_n \circ \Upsilon \circ \alpha(\nu^\chi)) = 0.$$

By [16, Proposition 3.3],

$$\Upsilon \circ \alpha(\nu^\chi) = \varpi_{\mathbb{R}_+^\mathbb{N}}(\nu^\chi).$$

Therefore, expression (8.6) converges to zero as ℓ tends to infinity. Now we let also n tend to infinity. Then we can deduce that

$$d_P(\nu^{\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N}, \nu^{\tilde{\chi}_{\lfloor c_N^{-1}t \rfloor}^N})$$

converges to zero in probability as N tends to infinity.

As an immediate consequence, for $t_1, \dots, t_k \in \mathbb{R}_+$, the vector

$$(d_P(\nu^{\hat{\chi}_{\lfloor c_N^{-1}t_i \rfloor}^N}, \nu^{\tilde{\chi}_{\lfloor c_N^{-1}t_i \rfloor}^N}), i \in [k])$$

converges to zero in probability as N tends to infinity. Using Proposition 3.8 again, Slutsky's theorem, and that $(\hat{\chi}_s, s \in \mathbb{R}_+)$ has a.s. no fixed times of discontinuity, we deduce the convergence in distribution

$$(\nu^{\hat{\chi}_{\lfloor c_N^{-1}t_i \rfloor}^N}, i \in [k]) \xrightarrow{d} (\nu^{\hat{\chi}_{t_i}}, i \in [k]) \quad (N \rightarrow \infty).$$

Hence,

$$(\hat{\chi}_{\lfloor c_N^{-1}t_i \rfloor}^N, i \in [k]) \xrightarrow{d} (\hat{\chi}_{t_i}, i \in [k]) \quad (N \rightarrow \infty)$$

in the marked Gromov-weak topology. As this convergence determines the finite-dimensional distributions of possible limit processes, it now suffices to show relative compactness of the sequence of processes $((\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+), N \in \mathbb{N})$, see [11, Theorem 3.7.8]. \square

To show the desired relative compactness in the proof of Theorem 3.5, we use the following lemma.

Lemma 8.3. *Let $k \in \mathbb{N}_0$ and $N \geq 2$. Then,*

$$d_{\text{mGP}}(\hat{\chi}_k^N, \hat{\chi}_{k+1}^N) \leq 2N^{-1}(N - \#\pi_{k+1}^N) + c_N.$$

The bound in Lemma 8.3 has the following meaning. There are at most $N - \#\pi_{k+1}^N$ many individuals in generation k that have more than one offspring in generation $k+1$. For each such offspring, the associated external branch has length c_N which needs not coincide with the external branch length of the ancestor in generation k . There are also $N - \#\pi_{k+1}^N$ individuals in generation k that die, and each such death can drastically increase an external branch length in generation $k+1$ (this is the freeing phenomenon mentioned in Section 3.5). For the other individuals in generation k , the external branch lengths increase by c_N from generation k to $k+1$. As each individual has weight N^{-1} , the bound is a consequence of the definition of the marked Gromov-Prohorov metric and the coupling characterization of the Prohorov metric.

Proof. Let $L \subset [N]$ denote the set of the labels of the individuals of generation k that have offspring in generation $k+1$, that is,

$$L = \{i \in [N] : \exists j \in [N] \text{ with } A_k(k+1, j) = i\}.$$

By definition of the population model in Section 3,

$$(8.7) \quad \#L = \#\pi_{k+1}^N.$$

For all $j_1, j_2 \in [N]$ with $j_1 \neq j_2$ and $i_1 = A_k(k+1, j_1)$, $i_2 = A_k(k+1, j_2)$, by definition of the population model in Section 3,

$$(8.8) \quad \rho_{k+1}^N(j_1, j_2) = \rho_k^N(i_1, i_2) + 2c_N.$$

For $i \in [N]$, we define the set

$$C_i = \{j \in [N] \setminus \{i\} : \rho_k^N(i, j) = \min\{\rho_k^N(i, \ell) : \ell \in [N] \setminus \{i\}\}\}.$$

In words, C_i consists of the individuals other than i with minimal distance to the individual i . That is, the set $C_i \cup \{i\}$ is the minimal clade of the individual i in the sense of [4]. Moreover, we define

$$M = \{i \in [N] : C_i \cap L \neq \emptyset, \exists! j \in [N] \text{ with } A_k(k+1, j) = i\}.$$

For $i \in M$, the individual i of generation k has exactly one offspring j in generation $k+1$, and at least one other member of the minimal clade of i has offspring in generation $k+1$. Hence, the minimal clade of i in generation k and the minimal clade of j in generation $k+1$ have the same most recent common ancestor. This implies, for i and j as above,

$$(8.9) \quad \min_{\ell \in [N] \setminus \{j\}} \rho_{k+1}^N(j, \ell) = \min_{\ell \in [N] \setminus \{i\}} \rho_k^N(i, \ell) + 2c_N.$$

We write $(r, v) = \beta(\rho_k^N)$ and $(r', v') = \beta(\rho_{k+1}^N)$. For $i \in M$, let $d(i)$ denote the label of the unique descendant in generation $k+1$ of the individual i of generation k . For all $i \in M$ and $j = d(i)$,

$$(8.10) \quad v'(j) = \frac{1}{2} \min_{\ell \in [N] \setminus \{j\}} \rho_{k+1}^N(j, \ell) = \frac{1}{2} \min_{\ell \in [N] \setminus \{i\}} \rho_k^N(i, \ell) + c_N = v(i) + c_N$$

by equation (8.9). For $i_1, i_2 \in M$ with $i_1 \neq i_2$ and $j_1 = d(i_1)$, $j_2 = d(i_2)$, it holds $j_1 \neq j_2$, and by equations (8.8) and (8.10)

$$(8.11) \quad r'(j_1, j_2) = \rho_{k+1}^N(j_1, j_2) - v'(j_1) - v'(j_2) = \rho_k^N(i_1, i_2) - v(i_1) - v(i_2) = r(i_1, i_2).$$

We define a relation \mathfrak{R} between the semi-metric spaces $([N], r)$ and $([N], r')$ by

$$\mathfrak{R} = \{(i, d(i)) \in [N]^2 : i \in M\}.$$

Equation (8.11) implies that the distortion of \mathfrak{R} equals zero,

$$\text{dis } \mathfrak{R} = \max\{|r(i_1, i_2) - r'(j_1, j_2)| : (i_1, j_1), (i_2, j_2) \in \mathfrak{R}\} = 0.$$

We set

$$\hat{\mathfrak{R}} = \{((i, v(i)), (j, v'(j))) \in ([N] \times \mathbb{R}_+)^2 : (i, j) \in \mathfrak{R}\}.$$

There exists a coupling ν of the probability measures $N^{-1} \sum_{i=1}^N \delta_{(i, v(i))}$ and $N^{-1} \sum_{j=1}^N \delta_{(j, v'(j))}$ on $[N] \times \mathbb{R}_+$ with

$$\nu(\hat{\mathfrak{R}}) \geq N^{-1} \#M = 1 - N^{-1}(N - \#M).$$

By equation (8.10), it holds $|v(i) - v'(j)| \leq c_N$ for $(i, j) \in \mathfrak{R}$. By Proposition 7.3, which also holds for marked semi-metric measure spaces, it follows

$$d_{\text{mGP}}(\hat{\chi}_k^N, \hat{\chi}_{k+1}^N) \leq N^{-1}(N - \#M) + c_N.$$

It remains to show

$$(8.12) \quad N - \#M \leq 2(N - \#L).$$

The assertion then follows by equation (8.7).

For $i \in [N]$, let $I_i = \mathbf{1}\{i \in L, C_i \subset L^c\}$. (Then I_i is the indicator variable that individual i reproduces as the only individual of its minimal clade.) Let $i, j \in [N]$ with $i \neq j$ and consider the case that there exists $\ell \in C_i \cap C_j$. W.l.o.g., we assume $\rho_k^N(j, \ell) \leq \rho_k^N(i, \ell)$ (if this does not hold, we transpose i and j). As $\rho_k^N \in \mathfrak{U}$, we obtain

$$\rho_k^N(i, j) \leq \rho_k^N(i, \ell) \vee \rho_k^N(j, \ell) = \rho_k^N(i, \ell).$$

As $\ell \in C_i$, it follows $j \in C_i$. If $I_i = 1$, then it follows that $j \in L^c$ and $I_j = 0$. Hence, in any case, the elements of the set $\mathcal{A} := \{C_i : i \in [N], I_i = 1\}$ are nonempty disjoint subsets of L^c ,

or it holds $\mathcal{A} = \emptyset$. This implies $\#\{C_i : i \in [N], I_i = 1\} \leq N - \#L$. Furthermore, generation k contains at most $N - \#L$ many individuals with more than one offspring in generation $k + 1$. The claim (8.12) follows by definition of M . \square

Proof of Theorem 3.5 (end). We assume $\Xi \in \mathcal{M}_{\text{dust}}$. To show relative compactness of the sequence of processes $((\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+), N \in \mathbb{N})$, it suffices to verify condition (b) in Theorem 3.8.6 of [11]. Condition (a) in this theorem is satisfied as the one-dimensional distributions converge.

Using Lemma 8.3 and the bound

$$N - \#\pi \leq \#\cup \sigma$$

for $\pi \in \mathcal{P}_N$ and $\sigma = \{B \in \pi : \#B \geq 2\}$, we obtain

$$\mathbb{E}[d_{\text{mGP}}(\hat{\chi}_k^N, \hat{\chi}_{k+1}^N)] \leq 2N^{-1}\mathbb{E}[\#\cup \sigma_1^N] + c_N$$

for all $k \in \mathbb{N}_0$. By exchangeability,

$$\mathbb{E}[\#\cup \sigma_1^N] = \sum_{i=1}^N \mathbb{E}[\mathbf{1}\{i \in \cup \sigma_1^N\}] = N\mathbb{P}(\gamma_1(\sigma_1^N) = \{\{1\}\}).$$

Let $(\mathcal{F}_t^N, t \in \mathbb{R}_+)$ be the filtration induced by the process $(\hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N, t \in \mathbb{R}_+)$. For $t \in \mathbb{R}_+$, $\delta > 0$, $u \in [0, \delta]$, and $s \in [0, \delta \wedge t]$, the Markov property of $(\hat{\chi}_k^N, k \in \mathbb{N}_0)$ at $\lfloor c_N^{-1}t \rfloor$ yields

$$\begin{aligned} & \mathbb{E}[d_{\text{mGP}}(\hat{\chi}_{\lfloor c_N^{-1}(t+u) \rfloor}^N, \hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N) | \mathcal{F}_t^N] d_{\text{mGP}}(\hat{\chi}_{\lfloor c_N^{-1}(t-s) \rfloor}^N, \hat{\chi}_{\lfloor c_N^{-1}t \rfloor}^N) \\ & \leq \mathbf{1}\{\delta \geq c_N/2\}(\lfloor c_N^{-1}\delta \rfloor + 1)(2N^{-1}\mathbb{E}[\#\cup \sigma_1^N] + c_N) \\ (8.13) \quad & \leq \mathbf{1}\{\delta \geq c_N/2\}(\delta + c_N)(2c_N^{-1}\mathbb{P}(\gamma_1(\sigma_1^N) = \{\{1\}\}) + 1) \quad \text{a. s.} \end{aligned}$$

In the first inequality, we also use $d_{\text{mGP}} \leq 1$, and that if $\delta < c_N/2$, then at least one of the distances on the left-hand side of (8.13) equals zero.

Now we show that the right-hand side of (8.13) converges to zero uniformly in N as δ tends to zero. For each $\varepsilon > 0$, there exists $N_\varepsilon \geq 2$ such that for all $N \geq N_\varepsilon$, it holds

$$c_N^{-1}\mathbb{P}(\gamma_1(\sigma_1^N) = \{\{1\}\}) \leq 2\lambda_{1, \{\{1\}\}}$$

and $c_N < \varepsilon$. Hence the right-hand side of (8.13) is bounded from above by $(\delta + \varepsilon)(4\lambda_{1, \{\{1\}\}} + 1)$ for $N \geq N_\varepsilon$. For δ sufficiently small and $N \leq N_\varepsilon$, the right-hand side of (8.13) equals zero.

As the right-hand side of (8.13) does not depend on t , we have verified (8.28) and (8.29) in Theorem 3.8.6 of [11]. To verify also (8.30), hence condition (b) in Theorem 3.8.6 of [11], we estimate as above

$$\mathbb{E}[d_{\text{mGP}}(\hat{\chi}_{\lfloor c_N^{-1}\delta \rfloor}^N, \hat{\chi}_0^N)] \leq \lfloor c_N^{-1}\delta \rfloor (2N^{-1}\mathbb{E}[\#\cup \sigma_1^N] + c_N) \leq \delta(2c_N^{-1}\mathbb{P}(\gamma_1(\sigma_1^N) = \{\{1\}\}) + 1).$$

Also this expression converges to zero uniformly in N as δ tends to zero. \square

LIST OF NOTATION

Here we collect notation that is used globally in the article.

Miscellaneous

$\mathbb{R}_+ = [0, \infty)$, $\mathbb{N} = \{1, 2, 3, \dots\}$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $[N] = \{1, \dots, n\}$ for $N \in \mathbb{N}$

γ_n : restriction map in various contexts (p. 11, p. 15, p. 26)

d_P : Prohorov metric

(Marked) distance matrices

$\mathfrak{U}_N, \mathfrak{U}$: space of semi-ultrametrics on $[N]$, on \mathbb{N} (p. 3)

$\hat{\mathfrak{U}}_N, \hat{\mathfrak{U}}$: space of decomposed semi-ultrametrics on $[N]$, on \mathbb{N} (p. 3, p. 4)

α : retrieves the semi-ultrametric from a decomposed semi-ultrametric (p. 3, p. 4)

$\beta : \mathfrak{U}_N \rightarrow \hat{\mathfrak{U}}_N$: decomposition map into the external branches and the remaining subtree (p. 4)

$\Upsilon(\rho)$: vector of the lengths of the external branches in the coalescent tree associated with ρ (p. 3, p. 4)

(Marked) metric measure spaces

\mathbb{U}_N, \mathbb{U} : spaces of isomorphy classes of ultrametric measure spaces (p. 4)

$\hat{\mathbb{U}}_N, \hat{\mathbb{U}}$: spaces of isomorphy classes of marked metric measure spaces (p. 5)

$d_{\text{GP}}, d_{\text{mGP}}$: Gromov-Prohorov metric, marked Gromov-Prohorov metric (p. 4, p. 5)

ν^χ : distance matrix distribution of $\chi \in \mathbb{U}$, or marked distance matrix distribution of $\chi \in \hat{\mathbb{U}}$ (p. 4, p. 5)

$\nu^{N,\chi}$: N -distance matrix distribution of $\chi \in \mathbb{U}$, or N -marked distance matrix distribution of $\chi \in \hat{\mathbb{U}}$ (p. 5, p. 6)

\mathcal{U}^{erg} : space of distance matrix distributions (p. 5)

$\psi_N : \mathbb{U}_N \rightarrow \mathbb{U}_N, \hat{\psi}_N : \hat{\mathbb{U}}_N \rightarrow \hat{\mathbb{U}}_N$: construction of (marked) metric measure spaces (p. 4, p. 5)

$\alpha : \hat{\mathbb{U}}_N \rightarrow \mathbb{U}_N$: maps a decomposed unlabeled tree to an unlabeled tree (p. 21)

$\beta : \mathbb{U}_N \rightarrow \hat{\mathbb{U}}_N$: decomposes an unlabeled tree at the external branches (p. 9)

$\beta_0 : \mathbb{U}_N \rightarrow \hat{\mathbb{U}}_N$: adds the zero mark (p. 12)

$\mathcal{C}_n, \hat{\mathcal{C}}_n$: sets of bounded differentiable functions with bounded uniformly continuous derivative (p. 11)

Π : set of polynomials on \mathbb{U} (p. 11)

$\hat{\Pi}$: set of marked polynomials on $\hat{\mathbb{U}}$ (p. 11)

\mathcal{C} : a set of test functions on \mathcal{U}^{erg} (p. 11)

Partitions and semi-partitions

\mathcal{P}_N : Set of partitions of $[N]$, associated transformations (equation (3.5))

$\mathbf{0}_n = \{\{1\}, \dots, \{n\}\} \in \mathcal{P}_n$

$\#\pi$: number of blocks of a partition π

\mathcal{S}_n set of semi-partitions of $[n]$ (p. 10), associated transformations (p. 10)

$\Delta = \{x = (x(1), x(2), \dots) : x(1) \geq x(2) \geq \dots \geq 0, |x|_1 \leq 1\}$

$\Delta^N = \{x \in \Delta : |x|_1 = 1, Nx(i) \in \mathbb{N}_0 \text{ for all } i \in \mathbb{N}\}$

$\Delta_c = \{x \in \Delta : x(1) > c\}$

Genealogy in the Cannings model

$(x_k^N, k \in \mathbb{N})$: sequence in Δ^N that gives the family sizes (p. 6)

$(\pi_k^N, k \in \mathbb{N})$ sequence in \mathcal{P}_N that gives the families (p. 6)

$A_j(k, i)$: label of the ancestor in generation j of the individual i in generation k (p. 6)

$\rho_k^N(i, j)$: genealogical distance (p. 6)

c_N : pairwise coalescence probability (equation (3.1))

b_N : probability that a randomly sampled individual is in a non-singleton family (p. 8)

$\chi_k^N = \psi_N(\rho_k^N)$: unlabeled genealogical tree (equation (3.4))

$\hat{\chi}_k^N = \hat{\psi}_N(\beta(\rho_k^N))$: unlabeled genealogical trees, decomposed at the external branches (p. 8)

$(r_k^N, v_k^N), \tilde{\chi}_k^N = \hat{\psi}_N(r_k^N, v_k^N)$: another decomposition of the genealogical trees (p. 9, p. 10)

Tree-valued Fleming-Viot processes

$\mathcal{M}_1(\Delta), \mathcal{M}_{\text{dust}}, \mathcal{M}_{\text{nd}}$: Set of probability measures on Δ , subsets of the measures with and without dust (p. 7)

$\Xi = \Xi_0 + \Xi\{0\}\delta_0$ (equation (3.3))

$\lambda_\pi, \lambda_{n,\sigma}$: reproduction rates (p. 11, p. 12)

Acknowledgments. This work is part of my PhD thesis. Partial support from the DFG Priority Programme 1590 “Probabilistic Structures in Evolution” is acknowledged.

REFERENCES

- [1] J. Bertoin, *Random fragmentation and coagulation processes*, Cambridge University Press, Cambridge, 2006. MR2253162
- [2] M. Birkner, *Stochastic models from population biology*, 2005. Lecture notes for a course at TU Berlin. Available at http://www.staff.uni-mainz.de/birkner/lehre_archiv/smpb-30.6.05.pdf.
- [3] D. Blount and M. A. Kouritzin, *On convergence determining and separating classes of functions*, Stochastic Process. Appl. **120** (2010), no. 10. MR2673979

- [4] M. G. B. Blum and O. François, *Minimal clade size and external branch length under the neutral coalescent*, Adv. in Appl. Probab. **37** (2005), no. 3, 647–662. MR2156553
- [5] C. Cannings, *The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models*, Advances in Appl. Probability **6** (1974), 260–290. MR0343949
- [6] ———, *The latent roots of certain Markov chains arising in genetics: a new approach. II. Further haploid models*, Advances in Appl. Probability **7** (1975), 264–282. MR0371430
- [7] I. Dahmer and G. Kersting, *The total external length of the evolving Kingman coalescent*, Probab. Theory Related Fields (2016). <http://dx.doi.org/10.1007/s00440-016-0703-7>.
- [8] A. Depperschmidt, A. Greven, and P. Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Probab. **16** (2011), 174–188. MR2783338
- [9] ———, *Tree-valued Fleming-Viot dynamics with mutation and selection*, Ann. Appl. Probab. **22** (2012), no. 6, 2560–2615. MR3024977
- [10] A. Etheridge, *Some mathematical models from population genetics*, Lecture Notes in Mathematics, vol. 2012, Springer-Verlag, 2011. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009. MR2759587
- [11] S. N. Ethier and T. G. Kurtz, *Markov processes: Characterization and convergence*, Wiley, New York, 1986. MR838085
- [12] A. Greven, P. Pfaffelhuber, and A. Winter, *Convergence in distribution of random metric measure spaces (Λ -coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322. MR2520129
- [13] ———, *Tree-valued resampling dynamics Martingale problems and applications*, Probab. Theory Related Fields **155** (2013), no. 3-4, 789–838. MR3034793
- [14] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser, Boston, MA, 1999. MR1699320
- [15] S. Gufler, *Pathwise construction of tree-valued Fleming-Viot processes* (2016). arXiv:1404.3682.
- [16] ———, *A representation for exchangeable coalescent trees and generalized tree-valued Fleming-Viot processes* (2016). arXiv:1608.08074.
- [17] J. F. C. Kingman, *Exchangeability and the evolution of large populations*, Exchangeability in probability and statistics (Rome, 1981), 1982, pp. 97–112. MR675968
- [18] S. Kliem and W. Löhner, *Existence of mark functions in marked metric measure spaces*, Electron. J. Probab. **20** (2015), no. 73. MR3371432
- [19] W. Löhner, *Equivalence of Gromov-Prohorov- and Gromov’s \square_λ -metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17. MR3037215
- [20] G. Miermont, *Tessellations of random maps of arbitrary genus*, Ann. Sci. Éc. Norm. Supér. (4) **42** (2009), no. 5, 725–781. MR2571957
- [21] M. Möhle, *Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson-Dirichlet coalescent*, Stochastic Process. Appl. **120** (2010), no. 11, 2159–2173. MR2684740
- [22] M. Möhle and S. Sagitov, *A classification of coalescent processes for haploid exchangeable population models*, Ann. Probab. **29** (2001), no. 4, 1547–1562. MR1880231
- [23] J. Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. MR2245368
- [24] S. Sagitov, *Convergence to the coalescent with simultaneous multiple mergers*, J. Appl. Probab. **40** (2003), no. 4, 839–854. MR2012671
- [25] J. Schweinsberg, *Coalescents with simultaneous multiple collisions*, Electron. J. Probab. **5** (2000), no. 4. MR1781024
- [26] L. Stournaras, *The tree-valued Wright-Fisher model*, 2012. Diploma thesis, Universität Freiburg.
- [27] A. M. Vershik, *Random and universal metric spaces*, Dynamics and randomness II, Kluwer Acad. Publ., Dordrecht, 2004, pp. 199–228. MR2153330

GOETHE-UNIVERSITÄT, INSTITUT FÜR MATHEMATIK, POSTFACH 11 19 32, FACH 187, 60054 FRANKFURT AM MAIN, GERMANY
E-mail address: gufler@math.uni-frankfurt.de